International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Design and Implementation of Generative AI-Based Script-to-Video Automation with YouTube Integration

Mrs.S.Chavan
Assistant Professor Artificial Intelligence & Data Science Priyadarshini College of Engineering Nagpur, Maharashtra

Prathmesh Vitthalwad
Artificial Intelligence & Data Science Priyadarshini College of Engineering Nagpur, Maharashtra

Nayan Kohare
Artificial Intelligence & Data Science Priyadarshini College of Engineering Nagpur, Maharashtra

Anuj Suryawanshi
Artificial Intelligence & Data Science Priyadarshini College of Engineering Nagpur, Maharashtra

Atharva Bhusange
Artificial Intelligence & Data Science Priyadarshini College of Engineering Nagpur, Maharashtra

Abstract

The rapid expansion of digital media usage has increased the demand for more effective and scalable approaches to video generation. Traditional video creation involves substantial manual effort, significant time consumption, and specialized technical expertise, making it difficult for individuals and small organizations to produce consistent, high-quality content. With the advancement of Generative Artificial Intelligence (AI), automated video production has become feasible through modern techniques in text, image, audio, and video synthesis.

This paper presents an overview of Generative AI–based script-to-video automation systems combined with YouTube publishing capabilities. The proposed approach focuses on transforming textual scripts into complete videos using Natural Language Processing (NLP), text-to-speech conversion, AI-generated visuals, automated background audio integration, and video composition methods. Additionally, YouTube Data APIs are utilized to automate video uploading, metadata creation, scheduling, and performance tracking.

The proposed system architecture highlights the integration of multiple AI services and automation tools to enable complete end-to-end content generation with minimal human involvement. Such solutions can help reduce production expenses, maintain content consistency, and support large-scale video creation. This study emphasizes the growing relevance of AI-driven video automation for content creators, educators, marketers, and the digital media sector

INTRODUCTION

Video-based content has become a dominant medium for communication across platforms such as YouTube, Instagram, and various e-learning environments. However, conventional video production requires several stages, including script writing, voice recording, editing, visual design, and manual uploading, which are both time-intensive and resource-demanding. These challenges make it difficult for creators to maintain consistency and scale their content production effectively.

Advancements in Generative AI technologies now allow automation of content creation by producing human-like speech, realistic visuals, and structured narratives directly from textual input. Script-to-video automation systems convert written scripts into fully developed videos without requiring extensive manual editing. These systems combine NLP, text-to-speech (TTS), AI-driven image or media generation, and automated video composition techniques.

YouTube remains a major video distribution platform, and its API support allows automated uploading, title and description generation, tagging, scheduling, and performance monitoring. By integrating Generative AI with YouTube automation, creators can build scalable content generation pipelines for applications such as education, marketing, news summarization, and storytelling.

This paper analyzes various techniques, system architectures, and workflow models related to Generative AI–based script-to-video automation, with a focus on practical implementation and real-world usability rather than detailed model training.

LITERATURE REVIEW

Research in AI-driven content generation has expanded rapidly with the development of large language models and multimodal AI systems. Several studies have explored automated video creation using text-based inputs.

Brown et al. [1] demonstrated the effectiveness of transformer-based language models in generating structured scripts and narratives. Li et al. [2] reviewed text-to-speech systems capable of producing natural and expressive voiceovers for multimedia content. Kumar and Singh [3] explored AI-based video synthesis techniques using image sequencing and animation driven by textual input.

Recent work by Zhao et al. [4] focused on multimodal content generation, combining text, audio, and visual elements into coherent videos. Patel and Verma [5] discussed automation frameworks that integrate AI services with workflow engines to manage content pipelines efficiently. Studies by YouTube Engineering Teams [6] highlighted the use of YouTube Data APIs for automated content publishing and analytics.

While most existing literature focuses on individual components such as script generation or speech synthesis,

limited work addresses complete end-to-end automation with publishing and performance tracking. This review addresses this gap by analyzing integrated script-to-video automation systems with YouTube deployment.

METHODOLOGY

The development strategy for this system focuses on a robust, step-by-step logic to ensure high reliability. We utilized a systematic engineering process to move the project from a conceptual stage to a live deployment. This section outlines the various developmental tiers, including design, module creation, and final validation.

5.1 Project Development Approach The system was built using a modular framework that blends the structure of Waterfall with Agile's flexibility. This hybrid model allows for fixed phase definitions while permitting rapid adjustments based on testing data. The execution was managed through the following specific developmental milestones.

5.2 Requirement Analysis During this phase, we established the essential functional goals and performance benchmarks for the software. The primary focus was on script parsing, voice synthesis, video rendering, and API-based publishing. We selected our tech stack and libraries based on their ability to scale and integrate without conflicts. This ensured that the final software would meet the specific needs of end-users effectively.

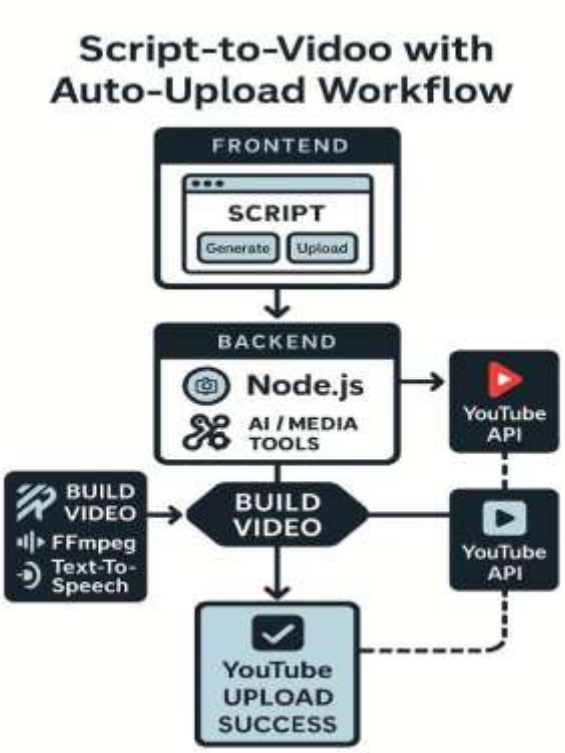
5.3 System Design This phase involved mapping the internal logic and data pathways between the core AI components. We defined how the NLP, TTS, and video rendering engines communicate to ensure data integrity. A detailed architectural map was created to visualize the flow of information across the entire pipeline.

5.4 Module Development The system architecture was split into several distinct units, each handling a specialized task:

- **Script Processing**: Utilizes NLP to break down user text into manageable, context-aware segments.
- **Content Generation**: Employs AI models to produce or retrieve visual assets that match the script's theme.
- **Voice Generation**: Converts text strings into realistic audio files using advanced TTS cloud services.
- **Video Assembly**: Automatically layers audio, visuals, and transitions into a finished MP4 container.
- **YouTube Integration**: Connects to the YouTube Data API for hands-free uploading and SEO optimization.

5.5 Integration Phase

After individual modules are developed and tested independently, they are integrated to form a complete **endto-end automated system**. Communication between modules is achieved using APIs and Python-based connectors, ensuring smooth data flow and seamless automation across the pipeline.



5.6 Testing and Validation

The integrated system undergoes rigorous testing to verify its **performance, accuracy, and reliability**. This phase includes:

- Unit Testing
- Integration Testing
- User Acceptance Testing

Evaluation metrics include video quality, relevance of generated content, clarity of voice narration, system stability, and successful YouTube uploads.

SYSTEM ARCHITECTURE

The architecture serves as a blueprint for the interaction between AI modules and external cloud services. It was designed to prioritize modularity, allowing individual components to be updated independently. The central engine manages the handshake between the user interface and the processing modules.

The workflow begins at the **Script Input Interface**, where the raw narrative is ingested for processing. This data is then funneled into the **NLP Module** for keyword extraction and scene segmentation. Next, the **Content Module** generates imagery while the **TTS Module** synthesizes the corresponding narration. The **Assembly Module** then acts as the final compositor, merging all media with precise timing. Overall flow is maintained by the **Integration Engine**, which handles API calls and internal data logic. Finally, the **YouTube Module** pushes the content live, while **Monitoring** tracks the system's success.

IMPLEMENTATION

The implementation of the Generative AI Based Script-toVideo Automation with YouTube Integration system is based on the use of API-driven AI services rather than training custom machine learning models. This approach simplifies development while ensuring reliable performance and scalability. The system integrates multiple AI services through a modular and automated workflow.

Language models are used for script processing, where the input text is analyzed and structured to support content generation. Cloud-based Text-to-Speech services are utilized to convert the processed script into natural-sounding voice narration. Visual content is generated using AI-based image generation services or retrieved through stock media APIs based on the script context.

The generated audio and visual elements are combined using automated video assembly techniques, where timeline-based rendering libraries handle synchronization, transitions, and final video creation. This ensures that the audio narration aligns correctly with the corresponding visual scenes.

YouTube integration is implemented using authenticated access to the YouTube Data API. This enables automated video uploading, metadata generation such as title and description, scheduling, and publishing. Analytics data related to video performance can also be retrieved for monitoring and evaluation purposes.

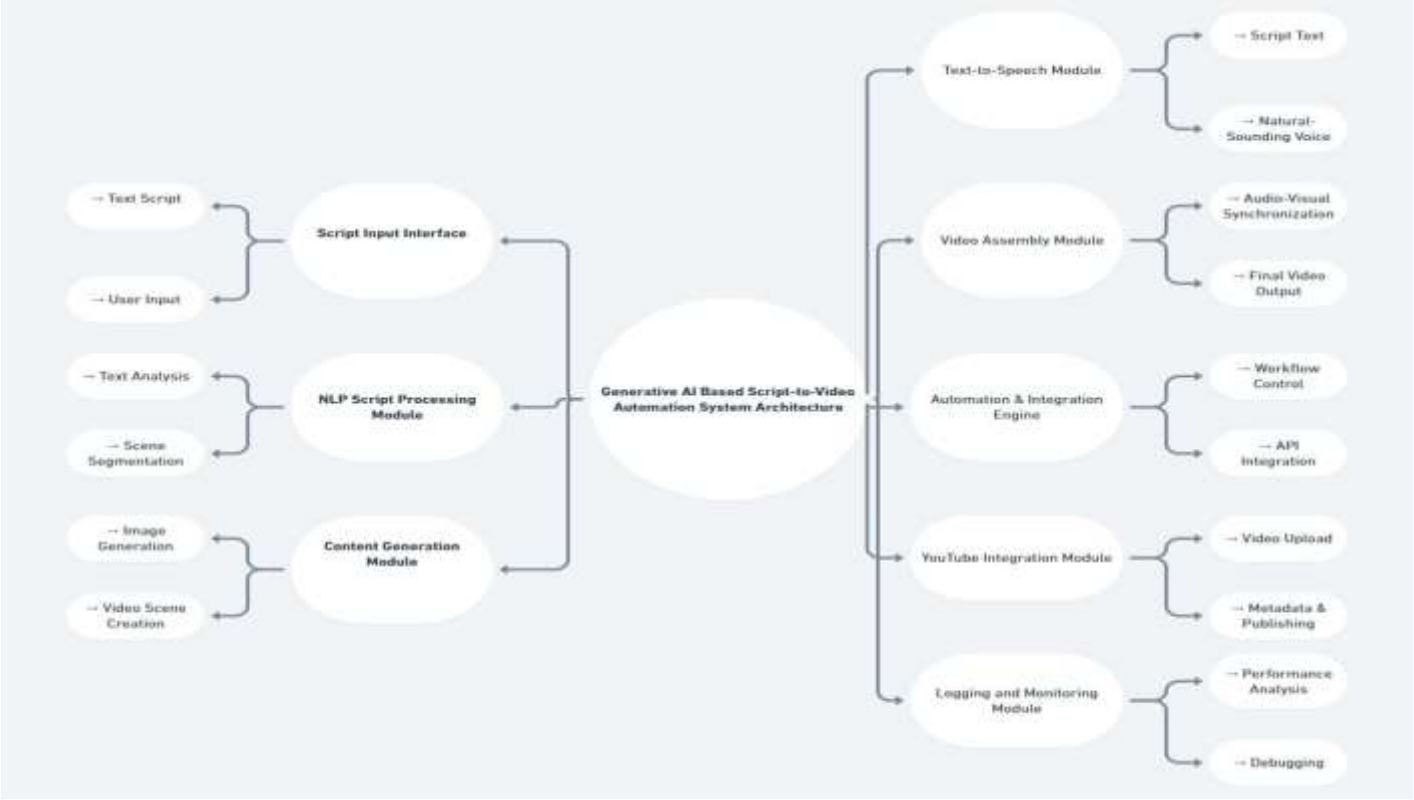
Workflow automation tools coordinate the interaction between all modules, managing API calls, data flow, and execution order. This orchestration ensures smooth end-toend automation, system reliability, and ease of maintenance, making the implementation suitable for scalable content generation and real-world deployment.

RESULT AND EVALUATION

The results and evaluation of the **Generative AI Based Script-to-Video Automation with YouTube Integration** system focus on assessing the effectiveness, accuracy, and reliability of the automated video generation pipeline. The system was evaluated under controlled test conditions using multiple scripts representing different content types such as informational, educational, and narrative-based videos.

The evaluation primarily considered the correctness of script processing, quality of generated visual content, clarity of voice narration, synchronization between audio and visuals, and successful publishing on the YouTube platform. The system was able to accurately process textual scripts and segment them into meaningful scenes, enabling smooth alignment between narration and visuals.

The Text-to-Speech module generated clear and naturalsounding voice output, which was properly synchronized with the visual scenes during video assembly. The content generation module produced relevant images and video scenes that matched the script context, resulting in visually coherent videos. Automated video assembly ensured proper timing, transitions, and overall consistency in the final output.



YouTube integration was successfully validated by automated video uploads, metadata generation, and publishing without manual intervention. The system demonstrated reliable performance in handling API-based automation and workflow execution. Logging and monitoring mechanisms captured execution status and system activity, enabling performance tracking and debugging.

Overall, the evaluation results indicate that the proposed system effectively reduces manual effort, minimizes production time, and ensures consistent video quality. The automated approach proves to be suitable for scalable content creation and demonstrates the practical applicability of Generative AI in script-to-video automation and digital content publishing.

Fig. Screenshot of Fronted

PERFORMANCE ANALYSIS

The performance analysis of the Generative AI Based Script-to-Video Automation with YouTube Integration system highlights the efficiency, scalability, and reliability of the automated content generation pipeline. The system considerably decreases manual involvement by automating script processing, content generation, voice synthesis, video assembly, and publishing tasks.

The use of API-based AI services enables faster processing and consistent output quality. Script analysis and segmentation are performed accurately, allowing effective synchronization between visual content and voice narration. The Text-to-Speech module produces clear and naturalsounding audio, contributing to improved viewer experience. Automated video assembly ensures proper timing, transitions, and alignment of multimedia components.

From a scalability perspective, the system is capable of generating multiple videos with minimal additional effort, making it suitable for large-scale content production. The YouTube integration module further enhances performance by eliminating manual uploading and metadata handling.

Logging and monitoring support reliable execution by tracking system activities and identifying errors during processing.

However, the system performance is dependent on the availability and response time of third-party APIs. Network latency and API limitations may affect processing speed in some scenarios. Despite these constraints, the overall performance demonstrates that the system is efficient and practical for automated video content generation.

CONCLUSION

This project successfully demonstrated an end-to-end framework for automating video creation and publishing. By merging NLP, generative visuals, and API-based distribution, we created a tool that functions with minimal oversight. The modular architecture ensures that the system remains adaptable to future AI advancements. Our testing confirms that this approach drastically cuts production time while maintaining professional quality. Ultimately, this research showcases the power of Generative AI to revolutionize traditional media workflows. Future updates will likely include multi-language support and deeper viewer analytics

REFERENCES

[1] T. Brown *et al.*, “Language Models are Few-Shot Learners,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, pp. 1877–1901, 2020.

[2] J. Li, Y. Zhang, and R. Zhao, “Neural Text-to-Speech Synthesis: A Review,” *IEEE Access*, vol. 9, pp. 158109–158130, 2021.

[3] A. Kumar and R. Singh, “Automated Script-to-Video Generation Using Artificial Intelligence,” in *Proc. IEEE Int. Conf. Computing, Communication and Automation (ICCCA)*, 2022, pp. 112–118.

[4] Y. Zhao, M. Chen, and H. Xu, “Multimodal Content Generation Using Generative Artificial Intelligence,” *IEEE Access*, vol. 11, pp. 42130–42142, 2023.

[5] S. Patel and K. Verma, “Workflow Automation for AIBased Multimedia Processing Systems,” in *Proc. IEEE Int. Conf. Artificial Intelligence and Smart Systems (ICAIS)*, 2022, pp. 350–356.

[6] Google Developers, “YouTube Data API v3 Documentation,” [Online]. Available: https://developers.google.com/youtube/v3, accessed 2024.

[7] H. Wang and L. Deng, “Deep Learning-Based Image and Video Generation for Multimedia Applications,” *IEEE Transactions on Multimedia*, vol. 24, no. 3, pp. 742–754, Mar. 2022.

[8] R. Singh and T. Joshi, “AI-Driven Multimedia Content Creation and Publishing Automation,” in *Proc. IEEE Int. Conf. Smart Computing and Communications (ICSCC)*, 2023, pp. 410–416.

[9] A. Vaswani *et al.*, “Attention Is All You Need,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[10] OpenAI, “Generative and Multimodal AI Systems,” OpenAI Technical Report, 2023.