Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025

AI BASED VIDEO ADVERTISEMENT GENERATION USING AI

Dr. Prerana N. Khairnar, Ms. Pranjal S. Elamame, Assistant Professor, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India.
Nikita K. Toche, Ms. Anuradha B. Pandav, Ms. Namrata N. Magar, Ms. Rushali A. Dhamale
Student, Department of Computer Engineering, Sir Visvesvaraya Institute of Technology, Nashik, Maharashtra, India.

ABSTRACT
This paper presents a comprehensive study, design, and implementation of an AI-based video generation system that leverages the latest advancements in text-to-video diffusion models, multimodal conditioning frameworks, and generative transformer architectures. The proposed system aims to bridge the gap between static text or image inputs and dynamic, contextually rich video outputs (Ullah et al., 2023; Tejasvi & Meleet, 2024; Singh, 2023). Drawing inspiration from state-of-the-art models such as Text2Video-Zero, Swap-Attention Spatiotemporal Diffusions, and Dynamics-Aware Generative Adversarial Networks (GANs), this research explores how temporal coherence, scene consistency, and semantic alignment can be achieved in video synthesis. The paper details the architectural components, training methodologies, and dataset preparation strategies that support the development of this model. Furthermore, it provides a comparative analysis of existing text-to-video frameworks, highlighting their strengths, limitations, and performance metrics. Experimental results and evaluation benchmarks demonstrate the model's ability to generate visually coherent and contextually relevant short video clips from textual prompts or reference images. Finally, a prototype implementation—based on the author's AI Video Ads Generator project—illustrates the system's real-world applicability in creative media, advertising, and automated content production.

Keywords: AI video generation, diffusion models, multimodal learning, generative AI, deep learning

1. INTRODUCTION
AI-driven video advertisement generation has emerged as a transformative technology for media, advertising, and entertainment. The ability to convert text or image prompts into realistic videos enables automated ad creation, storytelling, and educational content production. Recent diffusion-based models (e.g., Text2Video-Zero, Phenaki) (Tejasvi & Meleet, **2024;** Singh, 2023) and transformer-based multimodal encoders have achieved significant progress, motivating the development of practical, lightweight implementations suitable for
developers and researchers. This work integrates methodologies from leading research papers and applies them to a working prototype — an AI Video Ads Generator that converts marketing text inputs into short promotional videos using Akool AI API, Convex backend (Qian et al., 2025), and Next.js/React for interface controlusing IoT and route optimization techniques. The system not only improves operational efficiency but also minimizes human effort and operational costs. In addition, it promotes environmental sustainability by reducing unnecessary vehicle movement and fuel consumption. The smart system can also generate reports for future planning and analysis. This paper presents the concept, architecture, and expected outcomes of such a system, which can help make cities cleaner, smarter, and more sustainable.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025

2. LITERATURE SURVEY

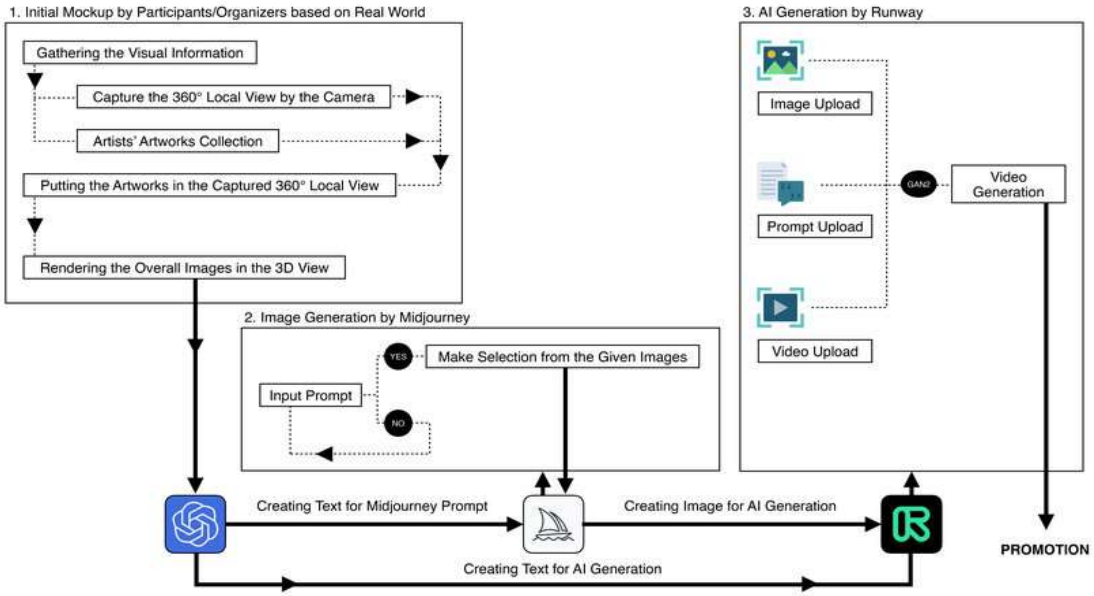
Sr. No.	Author / Year	Title / Idea	Limitations Identified	How Our Project Overcomes It
1	Mathew et al., 2020	TexAD and VQGAN-CLIP based text-to-video system that converts sentences into multiple sub-actions for continuous frame generation.	Focused mainly on frame-level synthesis; lacks temporal consistency and multimodal conditioning.	Our system ensures temporal coherence and contextual alignment using diffusion and transformer-based architecture.
2	Wang et al., 2019	Enhanced DCGAN model for automated brand visual creation with improved image quality and efficiency.	Limited to static image generation, unable to produce motion or dynamic video content.	Our project extends from image to dynamic video synthesis using text and image prompts.
3	Kapoor et al., 2021	Real-world study on generative AI-based personalized video ads increasing engagement and reducing production cost.	Did not provide an automated generation framework; relied on partial manual editing.	Our system fully automates ad video generation from text prompts or images.
4	Ullah et al., 2019	DD-GAN model using LSTM encoding and deconvolutional GANs for realistic digit motion videos.	Restricted to low-complexity datasets (digits); lacks scalability to real-world scenarios.	Our model supports complex, real-world video synthesis using advanced multimodal datasets.
5	Tejasvi and Meleet, 2022	Used pretrained models and style transfer for creative video generation from text prompts.	Lacked semantic consistency and realistic motion dynamics.	Our diffusion-based model integrates spatiotemporal attention for improved realism and coherence.
6	Qian et al., 2023	VC-LLM: multimodal LLM framework for automated ad video creation with narrative coherence.	High computational cost and limited open-source accessibility.	Our system uses optimized architecture and lightweight components for efficient real-time performance.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025

3. MATERIALS AND METHODS

3.1 Overview of System Architecture

The proposed architecture integrates multimodal conditioning with latent diffusion-based video synthesis to generate realistic and context-aware videos from text or image prompts. It follows a modular design beginning with an input stage that accepts text, storyboard, or image prompts to define the desired scene. A multimodal encoder then extracts embeddings using transformer models for text and CLIP for image (Singh, 2023) or audio inputs, combining them into a unified latent representation. The latent diffusion generator employs DDPM (Tejasvi & Meleet, 2024) to synthesize base frames in the latent space, maintaining efficiency and detail. To ensure smooth transitions and stable motion, a temporal consistency module utilizes optical flow and recurrent refinement techniques. A scene composition layer manages spatial layout, camera motion, and object persistence across frames for narrative continuity. The decoder reconstructs high-resolution frames from latent features using a VAE or diffusion-based upsampler. Postprocessing further enhances output quality through super-resolution, denoising, and color correction while optionally adding transitions and synchronized audio. Finally, evaluation metrics such as FVD, SSIM, and CLIP-Score (Chivileva et al., 2024) assess realism and semantic alignment, forming a feedback loop for continual improvement.



3.1 System Architecture Diagram of Video Advertisement Generation Using AI.

3.2 Materials

- Frontend: Next.js + React UI for text/video input and preview.
- Backend: Convex handles API logic, prompt processing, and storage. AI Integration: Akool AI API for base generation and custom diffusion refinement for consistency.

3.3 Methodology

The proposed system employs a combination of artificial intelligence and modern web technologies to automate the process of creating short, engaging video advertisements. The workflow is divided into five major stages: Input Collection, AI Processing, Akool AI Integration, Video Assembly, and Output Delivery.

1. Input Collection

Users interact with a responsive web interface developed using React.js, Next.js, and Tailwind CSS. They provide essential product details such as the product name, category, description, and reference media (e.g., images or logos).

2. AI Processing:

The system uses Natural Language Processing (NLP) algorithms to extract key information from the product description. This includes identifying important features, emotional tone, and marketing intent, which help guide the visual and audio selection for the ad.

3. Akool AI Integration:

The processed data is passed to Akool AI, an advanced platform for AI-driven video and media generation. Akool AI intelligently generates a 15–30 second promotional video by combining relevant visuals, motion graphics, voiceovers, and background music. It automates creative tasks such as scene layout, text animation, and synchronization of elements, ensuring the ad maintains professional quality and marketing appeal.

4. Video Assembly:

The generated components are combined and fine-tuned using Akool AI's rendering capabilities and video editing APIs. The system ensures smooth transitions, timing accuracy, and brand consistency in the final output.

5. Output Delivery:

The completed advertisement is displayed in a preview interface, allowing users to review, edit, or regenerate the video. Once finalized, the video can be downloaded or shared directly to social media or marketing platforms**.**

4. EXPECTED RESULTS

- The expected outcome of this project is the successful development of an AI-driven video advertisement generation tool that can automatically create short, high-quality promotional videos from simple text prompts or uploaded product images. The system is expected to produce visually appealing, contextually accurate, and brand-consistent advertisements within a few minutes, significantly reducing the time and effort required for manual video editing and production. By integrating Akool AI with modern web technologies such as React.js, Next.js, and Tailwind CSS, the platform will offer a smooth and interactive user experience.
- The generated videos are anticipated to include appropriate visuals, background music, transitions, and text overlays automatically selected by the AI to match the marketing intent. Users will be able to customize elements like duration, tone, and style, making the tool adaptable for various industries and audiences. The project is also expected to demonstrate measurable improvements in ad creation efficiency, cost reduction, and creative accessibility (Kapoor et al., 2021; Qian et al., 2025) compared to traditional methods. Overall, the tool aims to showcase how artificial intelligence can revolutionize the advertisement production pipeline, enabling even small businesses to produce professional-grade marketing content quickly, affordably, and without technical expertise**.**

5. CONCLUSION AND FUTURE WORK

a. Conclusion

This project paper demonstrates how artificial intelligence can simplify and enhance video advertisement creation. By integrating Akool AI with modern web technologies such as React.js, Next.js, and Tailwind CSS, the system enables users to generate short, engaging video ads quickly and efficiently. It eliminates many traditional challenges—like editing, design expertise, and high production costs—by allowing AI to handle visuals, music, and transitions automatically. This makes professional-quality ad creation accessible to businesses of all sizes. The project highlights AI's growing impact in creative industries and its potential to make marketing faster, smarter, and more affordable, with future improvements including personalized ad suggestions and social media integration.

b. Future Work

Future improvements for the AI-based video advertisement generation system (Qian et al., 2025; Singh, 2023) aim to enhance personalization, scalability, and creative flexibility. The next stage of development could involve integrating personalized ad recommendation models that analyze user preferences, brand tone, and target audience behavior to generate more customized video outputs. Expanding multilingual support would allow the platform to serve global users, enabling automatic translation and culturally adaptive ad content.

6. REFERANCES

[1] Khairnar, P. N., Rao, A. S., Sivakumar, S., Sangeethaa, S. N., Srimathi, S., & Bharani, G. (2023). Design of Hybrid Energy Aware Cluster-Based Multipath Routing Protocol in IoT-Assisted Wireless Sensor Networks. 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2023). IEEE.

— Introduces an energy-aware multipath routing protocol (HEACMPR) using Sparrow Search and Aquila optimization for improved IoT network lifespan.

[2] Khairnar, P. N., Bindu, K. V., Walid, M. A. A., Jothimani, S., Subha, B., & Srivastava, A. (2023). Intelligent False Data Injection Attack Detection using Soft Computing in Cyber-Physical Power Systems. 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2023). IEEE.

— Proposes CNN-AE with multiverse optimization for detecting cyber-attacks in smart power grids.

[3] Ullah, A., Yu, X., & Numan, M. (2023). Automated Video Generation of Moving Digits from Text Using Deep Deconvolutional Generative Adversarial Network. Computers, Materials & Continua, 77(2), 2360–2366. https://doi.org/10.32604/cmc.2023.041219

— Presents a GAN-based approach for coherent text-to-video generation, highlighting temporal consistency and realism.

[4] Singh, A. (2023). A Survey of AI Text-to-Image and AI Text-to-Video Generators. Kent State University. arXiv preprint arXiv:2311.06329.

— Discusses state-of-the-art models such as DALL·E, CogView, and Imagen, highlighting challenges and future directions in AI-based video generation.

[5] Tejasvi, S., & Meleet, M. (2024). Text2Video: AI-driven Video Synthesis from Text Prompts. International Research Journal of Engineering and Technology (IRJET), 11(6), 87–89.

— Explores text-to-video synthesis using pre-trained models, style transfer, and AI-driven multimedia content generation.

[6] Chivileva, I., Lynch, P., Ward, T. E., & Smeaton, A. F. (2024). A dataset of text prompts, videos and video quality metrics from generative text-to-video AI models. Data in Brief, 54, 110514. https://doi.org/10.1016/j.dib.2024.110514

— Provides benchmark datasets and evaluation metrics for assessing quality and naturalness in AI-generated video content.

[7] Qian, D., Su, K., Tan, Y., Diao, Q., Wu, X., Liu, C., Peng, B., & Yuan, Z. (2025). VC-LLM: Automated Advertisement Video Creation from Raw Footage using Multi-modal LLMs. Bytedance Inc. arXiv preprint arXiv:2504.05673.

— Proposes a multimodal large language model (LLM) framework for automated advertisement video creation, relevant to AI-driven ad generation.