

AI STORYBOARD GENERATOR FOR SHORT FILMMAKERS

¹Mrs. A. SRILATHA

Assistant Professor

²N. Vathsalika Reddy, ³Muneba Ifrah, ⁴G. Pravalika, ⁵K. Harshitha

Vathsalika2006@gmail.com, muneebaiifrah08@gmail.com, gpravali0206@gmail.com,

kondu.harshu@gmail.com

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING VIGNAN'S INSTITUTE OF MANAGEMENT AND TECHNOLOGY FOR WOMEN

(An Autonomous Institution)

(Affiliated to Jawaharlal Nehru Technological University Hyderabad, Accredited by NBA, NAAC with A+)

Kondapur(Village), Ghatkesar (Mandal), Medchal (Dist.)

Telangana-501301

(2025-2026)

ABSTRACT

The rapid advancement of artificial intelligence has significantly transformed creative industries, particularly in digital content creation and filmmaking. This project presents **CinePilot**, an AI-powered end-to-end storyboard and video generation platform designed to assist short filmmakers, students, and content creators in transforming simple ideas into complete cinematic outputs. The system enables users to input a one-line concept, which is then processed using advanced natural language processing and generative AI

techniques to produce structured stories, including scene breakdowns, character descriptions, dialogues, and narrative flow.

CinePilot further converts the generated story into a scene-wise storyboard and subsequently into short video clips using text-to-video generation models. To ensure continuity and coherence, the system incorporates a Visual DNA-based approach that maintains character and costume consistency across scenes. The generated video clips are stitched together using video processing tools to produce a final cinematic output. Additionally, the platform includes features such as character and costume design, teaser and trailer

generation, and a learning library containing filmmaking resources.

The proposed system addresses the limitations of existing tools by providing an integrated, automated, and cost-effective solution that reduces manual effort and technical complexity. By combining AI-driven automation with creative flexibility and collaborative features, CinePilot democratizes filmmaking and empowers users with minimal expertise to produce high-quality visual content efficiently.

Keywords

Artificial Intelligence, Storyboard Generation, Text-to-Video Synthesis, Natural Language Processing (NLP), Generative AI, Filmmaking Automation, Visual Storytelling, Scene Generation, Character Consistency, Video Stitching, Creative Content Generation, Human-Computer Interaction, Digital Media Production.

I. INTRODUCTION

The rapid evolution of Artificial Intelligence (AI) has significantly transformed various creative domains, including digital media and filmmaking. Traditionally, filmmaking is a complex and resource-intensive process that involves multiple stages such as script writing,

storyboarding, scene design, video production, and post-processing. These stages often require specialized skills, high production costs, and considerable time investment, making filmmaking less accessible to beginners, students, and independent creators.

In recent years, advancements in Natural Language Processing (NLP) and generative AI models have enabled machines to understand and generate human-like text, images, and videos. Several tools have emerged that assist in specific tasks such as script writing, storyboard creation, or video generation. However, most existing systems are fragmented and focus on individual components of the filmmaking pipeline, requiring users to manually integrate outputs from multiple platforms. This not only increases complexity but also leads to challenges in maintaining narrative consistency and character continuity across scenes.

To address these limitations, this project introduces **CinePilot**, an AI-powered end-to-end storyboard generator and filmmaking platform. The system is designed to transform a simple one-line idea into a complete cinematic output by automating key stages such as story generation, scene breakdown, storyboard creation, and video production. By

leveraging advanced AI techniques, CinePilot generates structured stories with dialogues, character descriptions, and visual prompts, which are further converted into scene-wise video clips.

A key feature of the proposed system is its ability to maintain consistency across scenes using techniques such as Visual DNA and prompt engineering. The platform also integrates video stitching tools to combine multiple clips into a final movie, along with additional functionalities such as character design, teaser generation, and a learning library for filmmaking resources. Furthermore, CinePilot incorporates a collaborative community environment that allows users to share projects and interact with other creators.

By combining automation, creativity, and accessibility, CinePilot aims to simplify the filmmaking process and make it more efficient and cost-effective. The system empowers users with minimal technical expertise to create high-quality cinematic content, thereby democratizing the field of digital storytelling and opening new opportunities for innovation in AI-driven filmmaking.

II. LITERATURE REVIEW

The integration of Artificial Intelligence (AI) in filmmaking and digital content

creation has gained significant attention in recent years. Various research efforts and tools have focused on automating different stages of the filmmaking pipeline, including story generation, storyboard creation, visual content synthesis, and video production. However, most of these approaches address specific components rather than providing a unified end-to-end solution.

One of the key areas of research is **Natural Language Processing (NLP) for story generation**. Transformer-based models, such as GPT architectures, have demonstrated strong capabilities in generating coherent and context-aware narratives from simple textual inputs. These models can produce structured stories with dialogues, character descriptions, and logical flow, making them suitable for scriptwriting tasks. Despite their effectiveness, they often face limitations in maintaining long-term coherence and creative diversity in extended narratives.

Another significant advancement is in **text-to-image generation**, where models like diffusion-based architectures are used to create high-quality images from textual descriptions. Tools such as DALL·E, Stable Diffusion, and MidJourney have been widely adopted for storyboard generation and concept art creation. While these

models produce visually appealing results, they struggle to maintain consistency in character appearance and scene continuity across multiple images, which is essential for storytelling applications.

In the domain of **text-to-video synthesis**, platforms like Runway Gen-4, Google Veo, and Pika Labs have introduced the ability to generate short video clips from textual prompts. These systems utilize advanced generative models to simulate motion and cinematic effects. However, they are typically limited to short-duration clips and often fail to ensure temporal consistency when generating longer sequences, making them less suitable for complete film production workflows.

Additionally, **AI-based storyboarding tools** such as Boords, Drawstory, and LTX Studio assist users in visualizing scripts by converting them into structured storyboards. These tools enhance pre-production planning but lack integration with video generation and post-production processes. Similarly, **video editing and processing tools** like FFmpeg provide efficient mechanisms for video stitching and rendering, yet they require manual intervention and do not incorporate intelligent content generation capabilities.

Overall, the literature indicates that while significant progress has been made in individual AI-driven components of filmmaking, there remains a gap in developing a comprehensive system that integrates story generation, storyboard creation, video synthesis, and editing into a single automated workflow. The proposed system, CinePilot, addresses this gap by combining multiple AI technologies into a unified platform, thereby improving efficiency, consistency, and accessibility in digital filmmaking.

III. METHODOLOGY

1. Input Acquisition and Preprocessing

The process begins with user input in the form of a one-line idea or prompt. Users can also specify parameters such as genre, style, and story length. The input is preprocessed to remove noise and structure it into a format suitable for AI-based story generation. This step ensures that the system understands the context and generates meaningful outputs.

2. Story Generation using NLP

The processed input is passed to advanced NLP models, such as transformer-based architectures, to generate a structured story. The generated content includes multiple components such as scene breakdowns,

character descriptions, dialogues, and narrative flow. Prompt engineering techniques are used to guide the model in producing coherent and context-aware stories. Additionally, a scene segmentation approach based on a multi-act structure ensures logical progression across the storyline.

3. Storyboard Generation

Once the story is generated, it is converted into a scene-wise storyboard. Each scene is transformed into a visual prompt describing the environment, characters, and actions. This step bridges the gap between textual storytelling and visual representation, enabling the system to prepare inputs for video generation models.

4. Video Generation (Text-to-Video Models)

The storyboard prompts are used as inputs to text-to-video generation models, such as diffusion-based AI systems. Each scene is converted into a short video clip (typically 5–10 seconds). These models generate realistic motion, lighting, and cinematic effects based on textual descriptions, forming the visual foundation of the final film.

5. Character Consistency using Visual DNA

To maintain uniformity across scenes, a Visual DNA technique is implemented. This method extracts key character attributes such as facial features, costumes, and traits, and embeds them into every prompt. By locking these attributes, the system ensures consistent character appearance and continuity throughout the generated video clips.

6. Audio Generation (Text-to-Speech)

Dialogue and narration generated during the story creation phase are converted into audio using Text-to-Speech (TTS) models. The system uses AI-based voice synthesis to produce natural-sounding speech, enhancing the realism of the generated video.

7. Video Stitching and Post-Processing

All generated video clips are combined into a single continuous video using video processing tools such as FFmpeg. This step includes video concatenation, audio synchronization, and optional background music integration. The final output is a complete cinematic video that aligns with the original user input.

8. Data Storage and Management

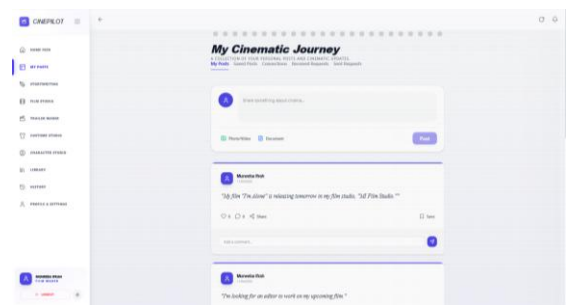
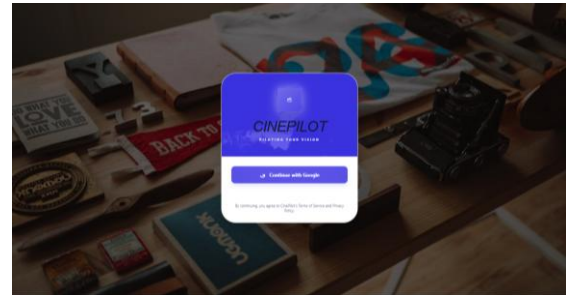
The system stores user inputs, generated stories, and video outputs in a database for

future access. This enables users to manage projects, revisit previous work, and make modifications if required.

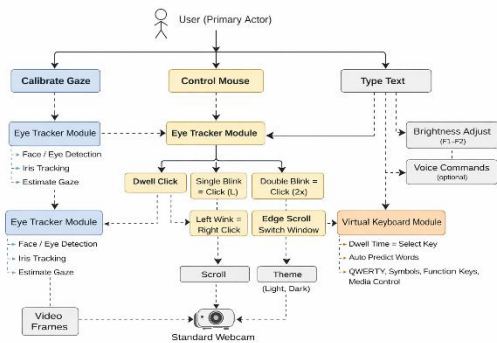
9. System Integration and Workflow Automation

All modules are integrated into a unified architecture consisting of frontend, backend, AI processing, and database layers. APIs are used to connect external AI services, and fallback mechanisms ensure continuous operation in case of service failures. This integration enables seamless automation of the entire filmmaking workflow.

and final cinematic video output, validating the feasibility of an end-to-end AI-driven filmmaking platform.



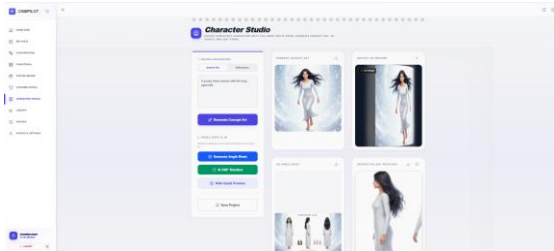
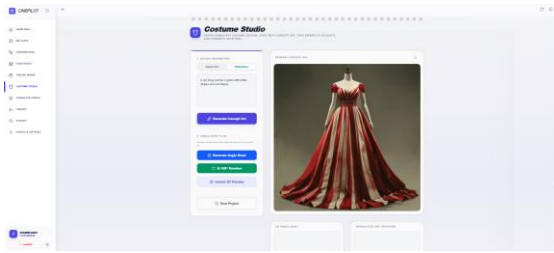
IV. SYSTEM ARCHITECTURE



V. RESULTS & DISCUSSION

The implementation of the proposed system, **CinePilot**, demonstrates the effectiveness of integrating Artificial Intelligence techniques for automating the filmmaking process. The system successfully converts a simple one-line user input into a structured story, storyboard,

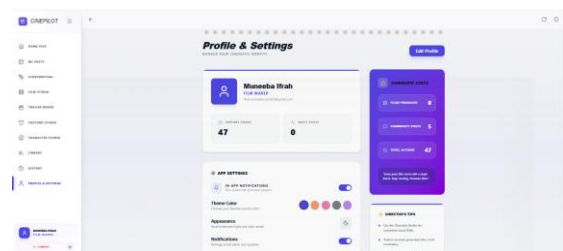
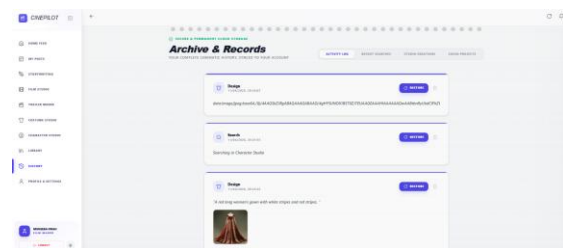
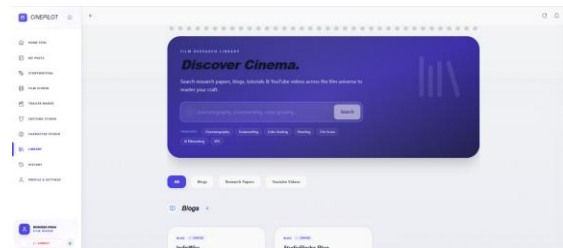
The **story generation module** produces coherent and context-aware narratives, including scene breakdowns, character descriptions, and dialogues. By utilizing advanced Natural Language Processing (NLP) models and prompt engineering techniques, the system maintains logical flow and narrative consistency across multiple scenes. The generated stories are adaptable based on user-defined parameters such as genre and style, demonstrating flexibility and personalization capabilities.



The **storyboard generation process** effectively transforms textual content into structured visual prompts, enabling clear visualization of scenes before video production. This step plays a crucial role in bridging the gap between textual storytelling and visual representation. The **text-to-video generation module** produces short video clips for each scene with reasonable visual quality, motion representation, and cinematic effects. However, minor variations in output quality are observed depending on the complexity of prompts and the performance of external AI APIs.

A significant contribution of the system is the implementation of **Visual DNA for character consistency**, which ensures that characters maintain similar appearances across different scenes. This improves narrative continuity compared to existing tools, although slight inconsistencies may

still occur due to limitations in current AI models. The **audio generation module** successfully converts dialogues into natural-sounding speech, enhancing the overall realism of the generated videos.



The **video stitching and post-processing module**, implemented using FFmpeg, effectively combines individual scene clips into a unified cinematic output. The synchronization of audio and video further improves the viewing experience. The system also demonstrates efficient data management by storing user projects, enabling reuse and modification of generated content.

Despite these strengths, certain limitations are observed. The system relies on third-party AI APIs, which may introduce latency, usage limits, or service interruptions. Additionally, generating high-quality video clips for multiple scenes can be time-consuming, and maintaining perfect visual consistency remains a challenge. These factors indicate areas for further optimization and improvement.

VI. CONCLUSION

This project presented **CinePilot**, an AI-powered end-to-end storyboard and filmmaking platform designed to simplify and automate the process of converting a simple idea into a complete cinematic output. The system successfully integrates multiple stages of filmmaking, including story generation, storyboard creation, video synthesis, and post-processing, into a unified and efficient workflow.

By leveraging advanced techniques such as Natural Language Processing (NLP), generative AI models, and video processing tools, CinePilot demonstrates the capability to generate structured narratives, maintain scene-wise continuity, and produce visually coherent video outputs. The incorporation of features like Visual DNA for character consistency, text-to-speech audio generation, and automated video stitching

enhances both the quality and realism of the final output. The system addresses key limitations of existing tools by providing an integrated, user-friendly, and cost-effective solution that reduces manual effort and technical complexity. It enables users with minimal expertise to participate in digital storytelling and filmmaking, thereby promoting creativity and accessibility. Additionally, the inclusion of learning resources and community features supports collaboration and skill development among users.

However, certain challenges remain, such as dependency on external AI APIs, variations in output quality, and processing time for generating video content. These limitations highlight the need for further improvements in model optimization, consistency handling, and system performance.

REFERENCES

1. T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997 – Used for understanding AI learning concepts and model behavior in generative systems.
2. Ashish Vaswani et al., “*Attention Is All You Need*”, NeurIPS, 2017 – Reference for Transformer architecture used in natural language processing for story generation.

3. Goodfellow et al., *Deep Learning*, MIT Press, 2016 – Used for understanding deep learning techniques applied in image and video generation models.
4. R. Socher et al., “*Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*”, EMNLP, 2013 – Reference for NLP-based structured text generation.
5. Runway ML Team, “*Gen-2 Text to Video Model Documentation*”, <https://runwayml.com> – Used as reference for AI-based text-to-video generation systems.
6. OpenAI, “*GPT Architecture and API Documentation*”, <https://platform.openai.com> – Used for understanding large language models in story generation.
7. Google DeepMind, “*Generative Video Models Research*”, <https://deepmind.google> – Reference for modern video synthesis and diffusion-based video generation techniques.
8. Fal.ai, “*AI Video Generation API Documentation*”, <https://fal.ai> – Used for implementing text-to-video generation pipeline in CinePilot.
9. Replicate Inc., “*Machine Learning Model Deployment Platform*”, <https://replicate.com> – Used for hosting and running diffusion models for image/video generation.