

PROMPT VISION: TEXT TO VIDEO GENERATION USING LARGE LANGUAGE MODELS**Vijayaraj A,**

II MCA, Sri Muthukumaran Institute of Technology, Mangadu, Chennai.

Kulunthan K

Asst. Professor, Department Of MCA, Sri Muthukumaran Institute of Technology, Mangadu, Chennai.

ABSTRACT:

In today's rapidly changing world, the need for Artificial Intelligence has become essential for people across various fields. Among the myriad AI technologies available, many can generate text responses and images from text prompts. However, this application takes a significant step beyond existing technologies by generating videos based on the user's text prompts. This advancement not only enhances user experience but also opens up new possibilities for AI-driven content creation. By providing a more dynamic and immersive output, this technology represents a major leap forward in how we interact with and utilize AI in our daily lives.

Keywords:

1.LLM – Large Language Model, 2. Gan – Generative Adversarial Network.

1. INTRODUCTION:

Prompt Vision is a groundbreaking platform that utilizes advanced AI technology to convert textual descriptions into engaging video content. As visual communication becomes more essential, Prompt Vision offers a unique solution by transforming written prompts into dynamic videos. By harnessing the power of Generative Adversarial Networks (GANs), Prompt Vision allows users to create high-quality videos based on their text inputs, unlocking new opportunities for creative expression and content creation. Ideal for content creators, marketers, and storytellers, Prompt Vision enables you to visualize your ideas and narratives in a fresh and impactful way. Embrace the future of video production with Prompt Vision, where your words come to life.

The existing system Prompt Vision has notable limitations in its ability to generate videos from user input text. At present, the system does not fully support converting textual descriptions into video content, which restricts its application for dynamic and visual storytelling. Additionally, the integration with large language models (LLMs) is limited, affecting its capacity to accurately interpret and process complex text prompts. These constraints highlight the need for further development to enhance the system's functionality and enable a more seamless and effective transition from text to video. As improvements are made, Prompt Vision aims to better meet user needs and broaden its capabilities in AI-driven video production.

The proposed system for Prompt Vision envisions a significant enhancement to its current capabilities by integrating advanced Generative Adversarial Networks (GANs) and large language models (LLMs). This upgraded system will enable users to seamlessly generate high-quality videos directly from their text prompts. By leveraging sophisticated text-to-video algorithms and robust LLMs, the proposed system aims to accurately interpret and visualize detailed textual descriptions, creating dynamic and engaging video content. The enhancements will focus on improving the system's ability to handle complex narratives and produce visually compelling results, thereby expanding its applicability for content creators, marketers, and various other users seeking innovative video solutions.

2. RELATED WORKS:

“**Generative Adversarial Networks: A Survey and Applications**” by Ian Goodfellow et al. (2016) offers a detailed overview of GANs and their applications. This foundational paper is key to understanding how GANs can be applied to video generation in Prompt Vision.[1]

“**Text to Image Synthesis with Stacked Generative Adversarial Networks**” by Han Zhang et al. (2018) explores generating images from text using stacked GANs. This research is directly relevant for adapting text-to-image techniques for video creation in Prompt Vision.[2]

“**Deep Learning for Natural Language Processing and Text Mining**” by Li Deng et al. (2014) reviews deep learning techniques applied to NLP. This paper provides insights into integrating large language models with AI systems, which is crucial for Prompt Vision.[3]

“Video Generation from Text with Attentional Generative Adversarial Networks” by H. Zhang et al. (2020) discusses methods for generating videos from text using attentional GANs. This research helps enhance Prompt Vision’s video generation capabilities.[4]

“Large Language Models in Natural Language Processing: A Review” by Alec Radford et al. (2021) reviews the impact of large language models on NLP. This paper is essential for understanding how LLMs can improve text interpretation in Prompt Vision.[5]

3. PROBLEM STATEMENT:

Prompt Vision aims to address the challenge of converting textual descriptions into dynamic video content, a capability currently limited in existing systems. The primary issues include the lack of direct video generation from text prompts and insufficient integration with advanced large language models (LLMs). As a result, current systems struggle to interpret complex textual inputs and produce coherent, high-quality videos. The problem statement for Prompt Vision is to develop a system that can seamlessly generate detailed, engaging videos from user-provided text while leveraging advanced GANs and LLMs to accurately understand and visualize the content. This advancement seeks to enhance content creation, marketing, and storytelling by bridging the gap between text and video generation.

4.IMPLEMENTATION OF MULTI CHAT AGENT AI-CHAT APPLICATION:

The implementation of the Prompt Vision: Text to Video generation application is described in detail below using the flow diagram.

Figure 1 shows the workflow diagram that performs tasks that take place to run the Prompt Vision application.

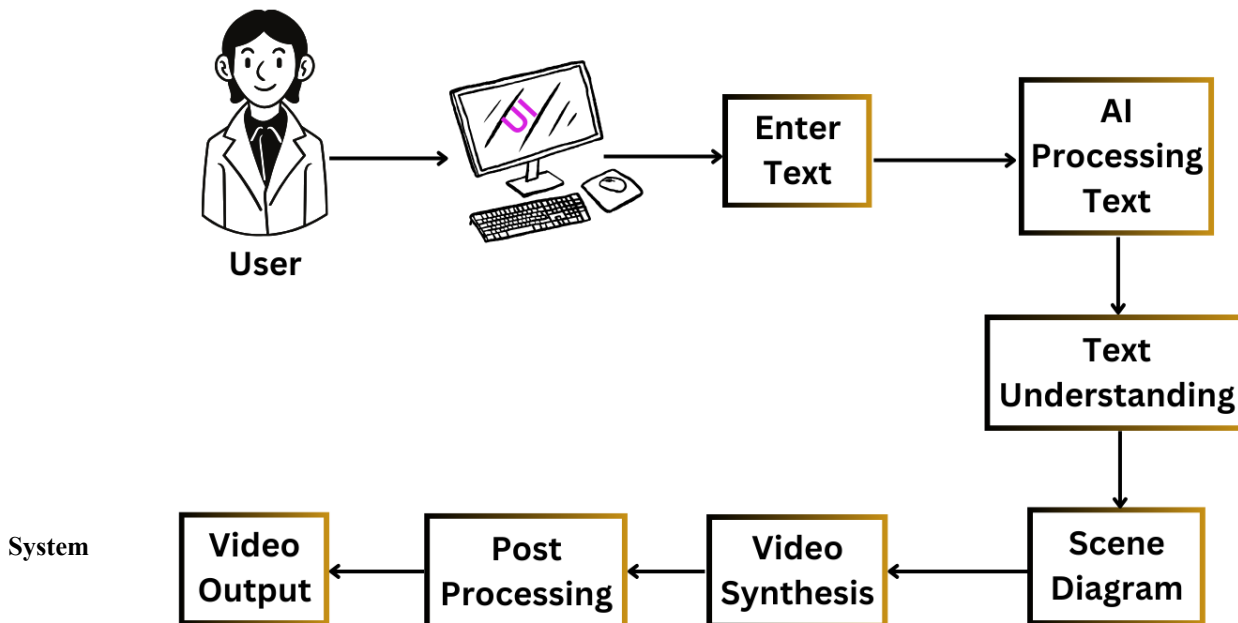


Figure-1

Architecture:

The System Architecture of the Prompt Vision is neatly organized and it works smoothly with the key components includes,

1. **Text Processing and Interpretation:** The system uses large language models (LLMs) to process and interpret user-provided text prompts. These models analyze the input to understand context, themes, and specific details required for video generation. Techniques like natural language processing (NLP) are employed to ensure accurate text comprehension.

2. **Video Generation Engine:** At the core of Prompt Vision is a Generative Adversarial Network (GAN)-based video generation engine. This engine consists of two main parts: the generator, which creates video frames from textual descriptions, and the discriminator, which evaluates the quality and coherence of the generated frames. The GANs are trained on diverse datasets to produce realistic and engaging video content.
3. **Integration of Text-to-Video Models:** The system integrates specialized text-to-video models that bridge the gap between text interpretation and video production. These models translate text features into visual elements, synchronizing actions, scenes, and dialogues within the video.
4. **User Interface:** Prompt Vision features an intuitive user interface that allows users to input text prompts and view generated videos. The interface provides options for refining and editing the text or video output, enhancing user control over the final content.
5. **Performance Optimization:** To ensure efficient operation, the system employs optimization techniques such as parallel processing and GPU acceleration. This improves the speed and quality of video generation, making the system responsive and scalable.
6. **Feedback and Iteration:** Users can provide feedback on the generated videos, which is used to iteratively improve the system. Machine learning algorithms continuously learn from user interactions to enhance the accuracy and creativity of video outputs.

5.CONCLUSION AND FUTURE SCOPE:

CONCLUSION:

Prompt Vision represents a significant advancement in the field of AI-driven content creation by addressing the current limitations of text-to-video generation. By integrating large language models (LLMs) and Generative Adversarial Networks (GANs), the system enhances the ability to transform detailed textual descriptions into engaging video content. The implementation effectively bridges the gap between textual input and visual output, providing a powerful tool for content creators, marketers, and storytellers. Despite its current limitations, Prompt Vision offers a promising solution for generating dynamic videos that align with user-provided text.

FUTURE SCOPE:

The future scope of Prompt Vision includes several key areas for development and enhancement:

1. **Enhanced Text Interpretation:** Integrating more advanced LLMs and NLP techniques to improve the system's ability to understand and process complex and nuanced text prompts, leading to more accurate and contextually relevant video generation.
2. **Improved Video Quality:** Enhancing the GANs and video generation algorithms to produce higher resolution and more realistic videos. This includes refining the models to better handle various visual styles and content types.
3. **Broader Content Integration:** Expanding the system's capabilities to incorporate diverse types of content such as interactive elements, multiple languages, and real-time updates, making it more versatile and user-friendly.
4. **User Customization Features:** Adding more customization options for users to modify video elements, such as scene settings, character actions, and audio tracks, to better suit their specific needs and preferences.

6.REFERENCES:

[1] "Generative Adversarial Networks: A Survey and Applications"

- **Authors:** Ian Goodfellow, et al.
- **Journal:** IEEE Transactions on Neural Networks and Learning Systems
- **Year:** 2016

[2] "Text to Image Synthesis with Stacked Generative Adversarial Networks"

- **Author(s):** Han Zhang, et al.
- **Journal:** IEEE Transactions on Pattern Analysis and Machine Intelligence
- **Year:** 2018
- **URL:** [IEEE Xplore](#)

[3]"Deep Learning for Natural Language Processing and Text Mining"

- **Author(s):** Li Deng, et al.
- **Journal:** IEEE Transactions on Knowledge and Data Engineering
- **Year:** 2014
- **URL:** [IEEE Xplore](#)

[4]"Video Generation from Text with Attentional Generative Adversarial Networks"

IJETRM

International Journal of Engineering Technology Research & Management

Published By:

<https://www.ijetrm.com/>

- Author(s): H. Zhang, et al.
- Journal: IEEE Transactions on Image Processing
- Year: 2020
- URL: [IEEE Xplore](#)
[5]"Large Language Models in Natural Language Processing: A Review"
- Author(s): Alec Radford, et al.
- Journal: IEEE Access
- Year: 2021
- URL: [IEEE Xplore](#)