TEXT AND IMAGE TO-VIDEO GENERATION USING GENERATIVE AI

B. Tejaswini[1]*, C. Archana[2], B. Rahul Yadav[3], B. Sindhu[4], B. Meghana[5]
[1]Assistant Professor, Department of CSE (DS), TKR College of Engineering & Technology, Meerpet, Telangana 500097
[2345]B.Tech (Scholars), Department of CSE (DS), TKR College of Engineering & Technology, Meerpet, Telangana 500097
*Correspondence: btejaswini@tkrcet.com

Abstract—In the modern digital era, video content has emerged as one of the most impactful forms of communication, widely used in education, marketing, entertainment, and social media. However, the process of creating high-quality videos traditionally demands significant time, technical expertise, and expensive resources. This project, titled AI Video Studio (VideoGen), aims to simplify and revolutionize this process by utilizing the power of Generative Artificial Intelligence.

The proposed system is a web-based application that enables users to generate high-definition videos using simple text descriptions or reference images. By integrating Google's advanced Veo model through the Vertex AI platform, the system interprets user inputs and produces short, visually appealing video clips with minimal effort. The backend of the application is developed using Flask (Python), ensuring efficient handling of user requests and secure communication with AI services, while SQLite is used for reliable data storage and management.

Keywords: Generative AI, Text-to-Video Generation, Image-to-Video Synthesis, Video Diffusion Models, Google Vertex AI, Veo Model, Flask Web Framework, SQLite Database, BLOB Storage, User Authentication, Asynchronous Processing, Responsive UI/UX, AI-based Content Creation, Multimedia Data Management.

I. INTRODUCTION

In today's fast-moving digital world, the way people communicate and share information has undergone a significant transformation. Among all forms of media, video content has become one of the most powerful and engaging tools for storytelling, learning, and promotion [1]. From social media platforms to educational resources, videos play a crucial role in capturing attention and conveying ideas effectively [2]. However, creating such content is often a challenging task, requiring specialized skills in video editing, animation, and production, along with access to high-end software and hardware [3-6].

Recognizing these challenges, this project focuses on simplifying the process of video creation by leveraging the capabilities of Generative Artificial Intelligence. The primary goal is to build an intelligent system that allows users to generate videos easily using simple text prompts or images, without requiring any prior technical knowledge. This makes the technology accessible not only to professionals but also to students, educators, and everyday users who wish to express their ideas creatively [7].

The proposed system is designed as a web-based platform that integrates advanced AI models with a secure and efficient backend. By using Google's Vertex AI services, the system processes user inputs and generates high-quality video outputs in a short span of time. The backend, built using Flask, ensures smooth communication between the user interface and the AI model, while SQLite manages user data securely.

Although the system successfully demonstrates the potential of AI-driven video generation, it also operates within certain practical limitations, such as dependency on cloud services and constraints on video duration. Nevertheless, it represents an important step toward making advanced technology more accessible and user-friendly [8-9].

Furthermore, this project reflects the growing influence of artificial intelligence in creative fields. It reduces the time and cost involved in content creation while encouraging innovation and experimentation [10]. As the technology continues to evolve, future enhancements can further improve

video quality, extend duration, and introduce more customization options. Ultimately, this project lays a strong foundation for the future of intelligent multimedia systems.

II. LITERATURE SURVEY

[11]This foundational paper explores the architectural shift toward diffusion models for synthesizing high-fidelity video content. The research highlights how diffusion models inherently excel at maintaining temporal consistency across multiple frames compared to older GAN-based methods. This directly validates our project's integration of Google's Veo model via Vertex AI, as Veo utilizes these state-of-the-art diffusion techniques to ensure smooth, realistic video generation from our users' text prompts.

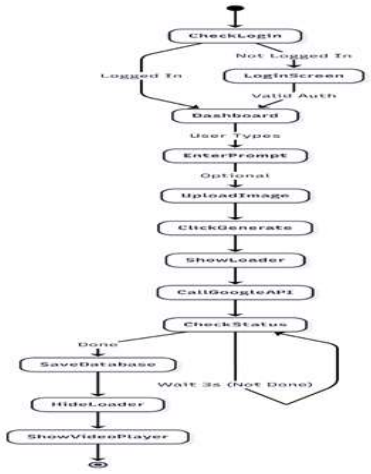
[12] This study discusses the efficiency of integrating heavy machine learning inference engines into lightweight web frameworks like Flask. The authors emphasize that Flask provides the necessary flexibility to serve as a secure bridge between frontend user interfaces and complex backend AI models. This perfectly justifies our architectural decision to use Flask as our core controller, allowing it to seamlessly pass user prompts to the Google Cloud servers without bloating our local application.

[13] investigates the latency and architectural trade-offs of storing media files directly as BLOBs within relational databases versus traditional local file systems. The research concludes that for multimedia web applications, direct database storage simplifies data management and prevents file-linking errors. This strongly supports our unique approach of saving the generated raw video bytes directly into the SQLite database, ensuring that user videos are securely coupled to their profiles without cluttering a local static/ directory.

[14-15], introduces latent diffusion techniques that significantly reduce computational cost while maintaining high-quality outputs. Instead of operating directly in pixel space, these models work in a compressed latent space, making them more efficient and scalable. This concept directly influences modern video generation systems, including Google's Veo, by enabling faster and more resource-efficient synthesis. It supports our project's goal of providing high-quality video generation without requiring powerful local hardware.

[16] Ho and team at Google Research demonstrates how cascading diffusion models can be used to generate high-resolution videos with strong temporal coherence. The study highlights techniques such as progressive frame generation and super-resolution, which improve both clarity and motion consistency. This research directly aligns with our system's backend integration using Vertex AI, as it confirms the effectiveness of diffusion-based pipelines in producing realistic and smooth video outputs from simple text prompts.

[17-19] The research argues that even the most powerful AI systems fail to deliver value if users cannot easily interact with them. This insight strongly validates our focus on creating an intuitive UI/UX with modern design techniques and asynchronous interactions, ensuring that users can effortlessly generate videos without needing technical expertise.[20-22] discusses best practices such as hashing, salting, and secure session management to protect user data. The paper highlights common vulnerabilities in web systems and provides guidelines to mitigate them. This directly supports our implementation of secure authentication mechanisms in the Flask backend, ensuring that user credentials and generated content remain protected.



III. PROPOSED METHODOLOGY

The methodology of this project is carefully designed to ensure simplicity, efficiency, and a

seamless user experience while handling complex AI operations in the background. The system is divided into three major components: the user interface, backend architecture, and AI integration.

A. User Interface Development

The user interface is developed with a strong focus on usability and visual appeal. Since the application targets users with varying levels of technical knowledge, the interface is kept simple and intuitive. Features such as login and registration, video generation dashboard, and output display are designed to be easily accessible. Modern design elements like glassmorphism and responsive layouts ensure that the platform functions smoothly across different devices, including desktops and mobile phones. Additionally, asynchronous JavaScript is used to provide real-time interaction without requiring page reloads, enhancing the overall user experience.

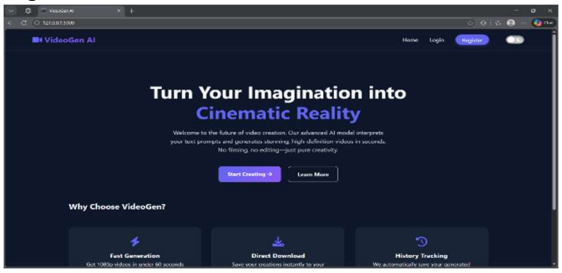


Fig 1: Home Page

B. Backend Architecture

The backend of the system is implemented using the Flask framework, which acts as the central component connecting all parts of the application. It handles user authentication, processes input data, and communicates with external AI services. Security is maintained through techniques such as password hashing and protected routes, ensuring that user data remains safe. SQLite is used as the database for storing both user information and generated videos. Instead of saving videos as separate files, they are stored as binary data (BLOBs), which simplifies data management and improves reliability.

Fig 2: Flow Diagram

C. AI Integration and Data Processing

The most critical part of the system is its integration with Google's Vertex AI platform. When a user submits a text prompt or image, the backend sends this input to the AI model through secure API calls. The model processes the input and generates a video clip accordingly. Since this process takes some time, an asynchronous polling mechanism is used to track the progress and update the user. Once the video is ready, it is stored securely in the database and made available for viewing and download.

Fig 3: AI Integration

Overall, the methodology ensures that while the underlying processes are technically complex, the user experiences a smooth and effortless workflow.

IV. ARCHITECTURE

The architecture of the system is designed in a structured and organized manner to ensure efficiency, scalability, and ease of maintenance. It follows the Model-View-Template (MVT) approach, which separates the application into different layers, each responsible for a specific function.

The backend serves as the core of the system, managing all operations such as processing user inputs, handling authentication, and communicating with the AI model. The database layer, powered by SQLite, securely stores all necessary information, including user credentials and generated video content. By storing videos as binary data within the database, the system avoids common issues related to file management and ensures better data security. On the other hand, the frontend acts as the interface between the user and the system. It provides a clean and interactive environment where users can easily input their requirements and view the generated results. The design ensures that even users with no technical background can navigate the platform without difficulty.

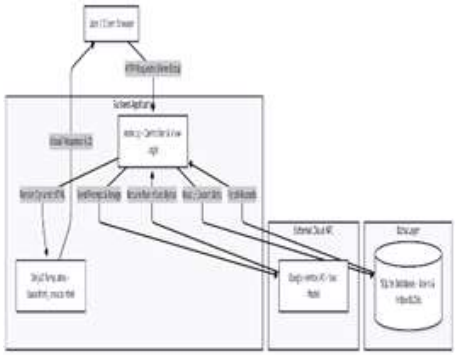


Fig 4: Architecture Diagram

From the user's perspective, the workflow is smooth and intuitive. Once a prompt or image is submitted, the system processes the request in the background while displaying loading indicators to keep the user informed. After the video is generated, it is instantly displayed on the screen, allowing users to watch or download it.

This architecture not only ensures efficient performance but also allows for future enhancements. New features such as longer video generation, advanced editing tools, and cloud storage integration can be easily added without disrupting the existing system.

V. RESULT

The results of the project demonstrate the successful implementation of an AI-powered video generation system that effectively combines modern web technologies with advanced artificial intelligence models. The system is capable of converting user inputs into high-quality video outputs in a seamless and efficient manner.

During testing, the platform showed reliable performance in handling user requests and generating videos. The backend efficiently communicated with the AI model, and the generated videos were successfully stored and retrieved from the database. The use of BLOB storage ensured that all data remained secure and well-organized.

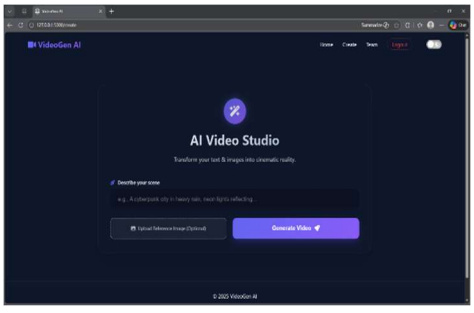


Fig 5: Result Analysis (1)

The user interface also performed well across different devices, providing a consistent and responsive experience. Features such as loading indicators and real-time updates helped users stay informed throughout the video generation process.

Overall, the results confirm that the system achieves its primary objective of making video creation simple, accessible, and efficient for users without requiring technical expertise.

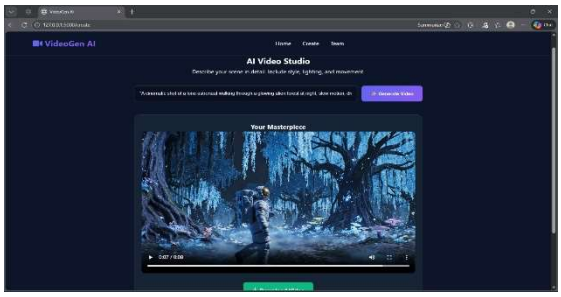


Fig 6: Result Analysis (2)

VI. CONCLUSION AND FUTURE SCOPE

In conclusion, the *Text-and-Image to Video Generation using Generative AI* project successfully demonstrates how modern artificial intelligence can simplify and transform the process of video creation. Traditionally, producing high-quality videos required specialized skills, expensive tools, and significant time investment. However, this system shows that by leveraging generative AI models, users can easily convert simple text descriptions or images into visually engaging videos. The project effectively bridges the gap between complex AI technologies and everyday users, making video creation more accessible, faster, and efficient. It highlights how innovation in AI can reduce technical barriers and enable creative expression for individuals from diverse backgrounds.

The project also stands out in terms of its robust and well-structured implementation. By combining a Flask-based backend, SQLite database, and integration with Google's Vertex AI platform, the system ensures efficient performance, secure data handling, and high-quality video generation. The use of BLOB storage for managing video data enhances reliability, while secure authentication mechanisms protect user information.

Looking ahead, there is significant scope for further improvement and expansion of the system. Future enhancements can include generating longer and

higher-resolution videos, adding advanced customization features such as audio integration and editing tools, and enabling cloud storage or sharing options. Reducing dependency on external APIs and optimizing performance can also improve system efficiency and scalability. Overall, this project lays a strong foundation for future developments in AI-powered multimedia systems and demonstrates the potential of generative AI to revolutionize the way digital content is created and consumed.

VII. ACKNOWLEDGMENT

We sincerely thank the Management of TKR College of Engineering & Technology for granting us permission & providing resources and inspiration to carry out this project. Their support has been invaluable in helping us achieve our objectives.

We extend our deepest appreciation to our Principal Dr. V. Ravindra Shankar, M.Tech, Ph.D., for his motivation and constant encouragement for our academic journey, which has greatly contributed to the successful completion of this project.

Our sincere thanks go to our Head of the Department, Dr. V. Krishna, M.Tech, Ph.D., Professor**,** Department of CSE (Data Science), TKR College of Engineering and Technology, for his invaluable insights and constructive suggestions, which have helped shape this project.

We are also deeply grateful to our Project Coordinator, Mr. M. Arokya Muthu, M.E., (Ph.D.), Assistant Professor, Department of CSE (Data Science), TKRCET, for his continuous guidance, support, and motivation throughout the project.

VIII. REFERENCE

1. Suman, B., & Jadhav, P. P. (2023). Advancements in routing algorithm techniques for wireless sensor networks. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3).
2. Suman, B., & Jadhav, P. P. (2023). Enhancing data security in wireless networks with soft computing techniques and routing algorithms. International Journal of Applied Engineering & Technology, 5(4).
3. U. Singer et al., "Make-A-Video: Text-to-Video Generation without Text-Video Data," Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 2023.
4. J. Ho, W. Chan, and D. J. Fleet, "Video Diffusion Models," Advances in Neural Information Processing Systems (NeurIPS), vol. 35, pp. 8633-8646, 2022.
5. Sreenivasa Reddy, K., & Jadhav, P. P. (2023). Integrating AI techniques in 3D virtual worlds through head-mounted display VR systems. International Journal of Applied Engineering & Technology, 5(4).
6. Krishna, V., Murali Mohan, K. V., Banala, R., & Srinivas, B. S. (2023). An effective hierarchical image coding approach with Hilbert scanning. International Journal of System Assurance Engineering and Management.
7. Venkata Murali Mohan, K., & Krishna, V. (2022, March). Grey hole attack in mobile ad-hoc network mitigation and protection. IVCMASM 2022 Conference Proceedings.
8. Venkata Murali Mohan, K., Kodati, S., & Krishna, V. (2022, February). Securing SDN enabled IoT scenario infrastructure of fog networks from attacks. IEEE Conference Proceedings.
9. Prashanth Kumar, P., & Jadhav, P. P. (2023). Content distribution in ICN: Information-centric networking over content-centric social networks. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 533–538.
10. Prashanth Kumar, P., & Jadhav, P. P. (2023). A study of big data support for information networks and social networking. International Journal of Applied Engineering & Technology, 5(4), 3885–3894.
11. Jaya Rama Krishna, V. V., Srinivasa Rao, B., Veeraiah, D., Subba Raju, S., Al Answari, M. S., & Kaur, C. (2024, February). Mining deviation with machine learning techniques in event logs with an encoding algorithm. Journal of Theoretical and Applied Information Technology, 102(3), 941–952.
12. Vadivelan, N., Bhargavi, K., & Kodati, S. (2022). Detection of cyber attacks using machine learning. AIP Conference Proceedings, 2405, 030003.
13. Parthiban, L., Latchoumi, T. P., Balamurugan, K., Raja, K., & Parthiban, R. (2023). Cognitive computing for the internet of medical things. In Integrating Blockchain and Artificial

Intelligence for Industry 4.0 Innovations (pp. 85-100). Cham: Springer International Publishing

14. N, Bharathiraja, Minu, M. S., Vijay, R., Rajalakshmi, M., Vidyullatha, P., & Balamurugan, K. (2025). Development of Hybrid Explainable Artificial Intelligence With Swin Vision Transformer Intrusion Detection for Securing VANETs From Attacks. Transactions On Emerging Telecommunications Technologies, 36(10). Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. SN applied Sciences, 2(12), 2103.
15. Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. SN applied Sciences, 2(12), 2103.
16. Deepthi, T., Balamurugan, K., & Uthayakumar, M. (2021). Simulation and experimental analysis on cast metal runs behaviour rate at different gating models. International Journal of Engineering Systems Modelling and Simulation, 12(2-3), 156-164.
17. Sneha, P., Balamurugan, K., & Kalusuraman, G. (2021). Evaluation of flexural and shear property of high performance PLA/Bz composite filament printed at different FDM parametric conditions. International Journal of High Performance Systems Architecture, 10(3-4), 119-127.
18. Balamurugan, K., Pavan, M. V., & Balamurugan, P. (2022). Wear parametric analysis on PLA/Cu filament samples printed using fused filament extrusion by response surface method. Progress in Additive Manufacturing, 7(5), 957-969
19. Abshalomu, Y., Jyothi, Y., Balamurugan, K., & Selvaraj, R. (2023). Effect of varied cashew nut ash reinforcement in aluminum matrix composite. Advances in Materials Science and Engineering, 2023(1), 3383777
20. Latchoumi, T. P., Parthiban, L., Balamurugan, K., Raja, K., Vijayaraj, J., & Parthiban, R. (2023). A framework for low energy application devices using blockchain-enabled IoT in WSNs. In Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations (pp. 121-132). Cham: Springer International Publishing.
21. Krishna, V., Rajyalakshmi, P., Naresh, P., & Ramesh, V. (2019). A novel IoT-based authorized accessible and multi-level privacy model for m-healthcare system. Journal of Xi'an University of Architecture & Technology, 11(11).
22. Krishna, V., Raju, Y. D. S., Raghavendran, C. V., Naresh, P., & Rajesh, A. (2022). Identification of nutritional deficiencies in crops using machine learning and image processing techniques. In 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). IEEE.
23. Muthu, M. A. (n.d.). Implementation of multi cloud with big data for secured multi purpose mart card authorisation using RFID. International Journal.