An AI-Based Platform for Automated Text, Image, and Video Content Generation

Prof. Rekha B.K[1], Sriee Amruthaa V[2], Shriya Hegde[3], Srushti Goudreddy[4], T. A. Keerthi[5]

[1]Professor, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India

[2,3,4,5] Student, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology, Bengaluru, Karnataka, India

***Abstract*—The increasing reliance on digital content across industries has accelerated the development of AI-driven tools for generating various forms of media. This project introduces an AI-based application capable of producing text, images, and videos from natural language prompts using advanced generative models. Leveraging Google Gemini for text, Replicate's Imagen-3 for image generation, and Tencent's Hunyuan-Video model for video synthesis, the system provides a unified solution for multimodal content creation. Built with Next.js and integrated with Clerk for authentication, Firebase for data management, and Google Cloud Storage for handling video files, the platform demonstrates how modern technologies can streamline content creation through automation and scalability.**

***Index Terms*—Generative AI, Multimodal Content Generation, Google Gemini, Imagen-3, Hunyuan-Video, Replicate API**

I. INTRODUCTION

The evolution of artificial intelligence has paved the way for tools that can autonomously generate high-quality digital media, addressing the increasing demand for rapid content creation in fields like marketing, education, and entertainment. Traditional content production often requires considerable manual effort and time. This project presents a modern AI-powered web platform that simplifies the creation of diverse content—including text, images, and videos—by utilizing prompt-based input. The system integrates several state-of-the-art technologies: Google Gemini for generating coherent and context-aware text, Replicate’s Imagen-3 model for high-resolution image creation, and Tencent’s Hunyuan-Video model for producing dynamic video clips from textual descriptions. The application also features Firebase for real-time database handling, Clerk for user authentication, and Google Cloud Storage for managing media assets. Developed using the Next.js framework, this platform represents a scalable and user-friendly approach to AI-assisted content generation.

II. RELATED WORK / LITERATURE REVIEW

From [1]. Recent advancements in artificial intelligence have introduced powerful tools like Google Gemini into the educational domain. Gemini, developed by DeepMind, is a next-generation multimodal AI model capable of processing and generating content across multiple formats, including text, images, audio, and video [1]. With its three-tiered structure—Gemini Nano, Pro, and Ultra—it caters to a wide range of users from mobile learners to advanced researchers. Its ability to deliver personalized learning experiences, support intelligent tutoring, and provide dynamic feedback marks a significant shift in educational technology. Gemini not only enhances engagement through interactive simulations but also assists in creating diversified educational content [1]. Despite its promising features, challenges such as ethical concerns, data privacy, and the lack of standardized educational guidelines remain. Overall, Gemini demonstrates the transformative potential of AI in education, provided its integration is guided by responsible and well-structured frameworks

From[2] This study developed an AI-based content generation platform integrating Google's Gemini language model to streamline digital content creation. The motivation stems from the increasing demand for efficient and creative content in digital domains such as marketing, education, and media. Traditional methods are often time-consuming and creatively limiting, prompting the need for automation using advanced AI. By integrating natural language processing (NLP) techniques with machine learning algorithms, the platform automates content creation while allowing customization in tone, style, and structure.The application architecture uses a Flask backend to handle user requests and process content through the Gemini API, which returns contextually relevant responses. The frontend, developed using JavaScript and styled with CSS, provides a responsive and user-friendly interface.Also emphasizes user experience (UX) through a responsive UI styled with CSS, and backend processing involving API key authentication, data parsing, and output formatting.The study also explores potential applications and acknowledges limitations like dependency on third-party APIs, data quality, and the need for continual updates based on user feedback and domain-specific customization.

From[3] This study of Google Gemini and other state-of-the-art text-generating AI models, including ChatGPT and GitHub Copilot. Emphasis is placed on the architectural design, learning algorithms, token handling capabilities, and multimodal integration of Gemini. The authors highlight the evolution of Google's AI—from Bard to the more capable Gemini Ultra and Gemini 1.5 models. Through empirical evaluation using metrics like BLEU, perplexity, and response generation time, the paper explores Gemini’s strengths in coherence, context retention, and efficiency. Key innovations include the Mixture-of-Experts architecture and a 1 million token context window. The paper also delves into Gemini’s integration with Google Workspace, multimodal inputs, and conversational memory features. Limitations such as data bias, factual inaccuracies, and contextual misunderstandings are acknowledged. Overall, the study underscores Gemini’s transformative potential in AI while recommending ethical and responsible development practices.

III. PROBLEM STATEMENT

In the current digital landscape, there is an ever-growing demand for efficient, user-friendly, and intelligent platforms that can support seamless content generation across multiple formats. Despite the availability of various machine learning models and AI-based tools for text, image, and video generation, most existing systems face significant limitations. These include a lack of unified multimodal integration, limited customization options, complicated user interfaces, and challenges in ensuring responsiveness and real-time generation.

Our project addresses these challenges by developing a web-based AI-powered content generation platform that leverages Google’s Gemini model in conjunction with third-party APIs such as Imagen-3 and Hunyuan-Video. The solution not only simplifies user interaction but also provides a tailored experience by allowing users to generate content aligned with specific styles, tones, and formats. By integrating these advanced models with tools like Firebase for storage and Clerk for authentication, the application ensures robust backend functionality while maintaining ease of use. This approach offers a significant improvement over traditional content generation systems, creating a pathway for personalized, real-time, and multimodal content creation.

IV. PROPOSED METHODOLOGY

Our project uses a modular and scalable architecture to facilitate the automated generation of multimodal content through a web-based application. At the core lies the integration of Google’s Gemini API, which provides advanced natural language processing capabilities. This section outlines the technological foundation and workflow of our system.

A. System Architecture The architecture consists of a frontend for user interaction, a backend for processing prompts, and third-party services for content generation and storage. Clerk is used for user authentication, while Firebase stores prompt data and content references. The Replicate platform acts as the gateway to AI models like Imagen-3 and Hunyuan-Video, enabling image and video generation respectively.

B. API Request and Response Flow When users submit a prompt, the frontend sends a POST request to the backend. The backend identifies the type of content requested (text, image, or video) and triggers the appropriate API call. Gemini is used for textual content, Imagen-3 for image generation, and Hunyuan-Video for video generation. Responses are parsed and formatted before being relayed to the frontend.

C. Gemini Integration for Text Generation The Gemini API is accessed through a secure key. When text generation is requested, the backend sends the prompt to Gemini, which processes the input and returns structured, context-aware content. This content is then formatted and displayed to the user.

D. Image and Video Content Handling The Replicate API facilitates visual content generation. Imagen-3 generates high-resolution images, while Hunyuan-Video processes short video clips from descriptive prompts. The content is stored in Google Cloud Storage, with links and metadata recorded in Firebase for future access.

E. User Experience and UI Components The frontend is developed using Next.js and styled with Tailwind CSS to ensure responsiveness. JavaScript ensures real-time interaction without page reloads. Users receive immediate visual feedback and can download or save the generated content.

V. IMPLEMENTATION

The implementation of our AI-powered content generation system is divided into frontend development, backend services, third-party API integration, and cloud infrastructure support.

A. Frontend Development Built using React and Next.js, the frontend offers a dynamic and intuitive interface. Clerk.js handles authentication, while the prompt form captures user input. Tailwind CSS styles the interface, ensuring responsiveness across all devices.

B. Backend and Middleware Services The backend, written in Node.js and TypeScript, handles logic flow and routes API requests. Middleware functions manage authentication tokens, request validation, and result formatting. Requests are made to Replicate APIs using Axios.

C. Replicate API Utilization For image generation, the google/imagen-3 model is used. A POST request with the user's prompt returns image URLs. Similarly, for video generation, the tencent/hunyuan-video model processes the input to generate dynamic videos. Response data is parsed and stored securely.

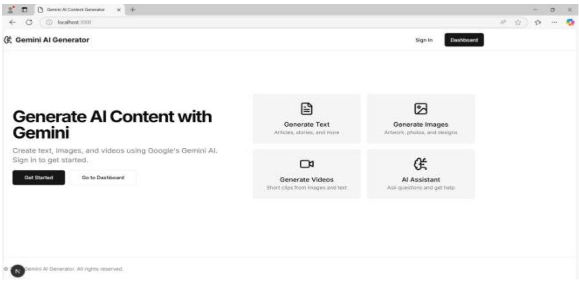
D. Cloud Infrastructure Firebase Firestore acts as the central data repository, storing user information and prompt history. Google Cloud Storage holds media files. Firebase SDK facilitates smooth interaction between services.

E. Output Presentation Once the content is generated, the backend sends it to the frontend, where it is presented in a formatted view. JSON structures are used for communication, and users are provided options to download or copy the results.

F. Tools and Libraries Key tools include React + Vite for rapid development, Axios for API communication, Tailwind for styling, Firebase SDK for cloud operations, and the Replicate SDK/REST APIs for AI model access. Together, these components form a powerful and modular content generation platform.

VI. RESULTS AND ANALYSIS

The system was tested with a variety of prompts ranging from descriptive narratives to simple commands. Text generation via Gemini yielded context-rich, coherent paragraphs. Imagen-3 produced high-resolution images with precise correspondence to prompts. Hunyuan-Video was successful in rendering short dynamic clips from brief textual inputs. The latency for generation varied: text (1-2s), images (3-6s), and videos (15-30s), depending on model load and API response. User feedback indicated ease of use, reliability, and potential applications in education, advertising, and content marketing.

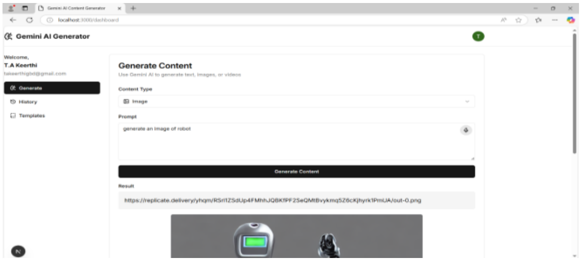

Gemini AI Content Generator
localhost:3000
Gemini AI Generator
Sign In
Dashboard
Get Started
Go to Dashboard
Generate Text
Articles, stories, and more
Generate Images
Artwork, photos, and designs
Generate Videos
Short clips from images and text
AI Assistant
Ask questions and get help

Generate AI Content with Gemini

Create text, images, and videos using Google's Gemini AI. Sign in to get started.

© Gemini AI Generator. All rights reserved.

Figure 1. Home page



Welcome,
T.A Keerthi
takeerthigbd@gmail.com

Generate Content

Use Gemini AI to generate text, images, or videos

Content Type: Image

Prompt: generate an image of robot

Result: https://replicate.delivery/yhqm/RSrl1ZSdUp4FMhhJQBKfPF2SeQMtBvykmq5Z6cKjhyrk1PmUA/out-0.png

Figure 2. Content generator

VII. CONCLUSION

In this project successfully demonstrates the potential of combining various AI models and web technologies to build a comprehensive multimodal content generation tool. By enabling users to input a simple text prompt and receive AI-generated text, images, or videos in return, the system addresses key challenges in digital content production. The integration of Clerk for user access, Firebase for backend support, and Replicate APIs for AI media generation highlights the practical feasibility of constructing powerful AI applications using existing cloud infrastructure. Overall, the platform offers a glimpse into the future of automated content creation, empowering users to transform ideas into multimedia content with minimal effort.

VIII. FUTURE WORK

To enhance the platform's capabilities and user experience, several improvements are proposed. One direction is the implementation of smart prompt enhancement using natural language processing to improve the quality of generated outputs. Support for additional languages and cultural context awareness could broaden the application's usability. Incorporating interactive image and video editing tools would allow for more control over generated media. A mobile version of the platform could increase accessibility and convenience. Moreover, migrating the AI models to private cloud infrastructure could help reduce API costs and improve system performance. Finally, adding collaboration features and version history can make the system more suitable for team-based creative projects.

REFERENCES

[1] M. Imran and N. Almusharraf, "Google Gemini as a next generation AI educational tool: a review of emerging educational technology," *Smart Learning Environments*, vol. 11, no. 22,2024.

[2] R. Shrivastav, S. Shahane, T. S. Hydri, M. V. Akre, and Z. D. Amin, "Exploring Potential of Gemini with AI Based Content Generator," *Int. J. Res. Comput. Inf. Technol.*, Special Issue, vol. 2, no. 1, pp. 68–72,2024

[3] Pande, R. Patil, R. Mukkemwar, R. K. Panchal, and S. B. Bhoite, "Comprehensive Study of Google Gemini and Text Generating Models: Understanding Capabilities and Performance," *Grenze Int. J. Eng. Technol.*, vol. 10, no. 2, pp. 857–863, Nov. 2024
Google Gemini

[4] https://deepmind.google/technologies/gemini/

[5] Replicate API: https://replicate.com/

[6] Imagen-3 by Google: https://imagen.research.google/

[7] Tencent Hunyuan-Video: https://replicate.com/tencent/hunyuan-video

[8] Firebase: https://firebase.google.com/

[9] Clerk Authentication: https://clerk.dev/

[10] Google Cloud Storage: https://cloud.google.com/storage