iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VI **Month of publication:** June 2025

DOI: https://doi.org/10.22214/ijraset.2025.71461

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

Genius AI A Unified Platform for Text, Image, Audio, Video, and Code AI

Sahil Meshram[1], Shivam Sahu[2], Shivam Kumar Gupta[3], Yatharth Sonteke[4], Nawaz Mehmood[5], Prof. Savita Sahu[6]
[1, 2, 3, 4, 5]*Student of Computer Science Engineering in Government Engineering College, Bilaspur*
[6]*Assistant Professor of Computer Science Engineering in Government Engineering College, Bilaspur*

Abstract: ***The rapid evolution of artificial intelligence (AI) has led to the development of specialized models across different modalities such as text, image, video, audio, and program code. This paper presents the design and conceptual framework for a multimodal AI platform that harmoniously brings together multiple AI systems into a single, user-friendly. The proposed platform leverages state-of-the-art AI models, each tailored for a specific modality—Natural Language Processing (NLP) models for text understanding and generation, Computer Vision models for image analysis and synthesis, Generative Video AI for dynamic scene creation, Audio AI for speech recognition and generation, and Code AI for intelligent code completion, debugging, and generation. This paper outlines the core design principles, technical challenges, system integration methods, and practical use cases, including educational tools and content creation. Our approach marks a significant step toward the realization of truly general-purpose AI platforms.***

Keywords: Multimodal AI, AI integration Platform, Text-to-Image, Audio-Visual AI, Code Generation AI, Smart Assistant Platform, Natural Language Processing (NLP).

I. INTRODUCTION

In recent years, artificial intelligence (AI) has made significant advancements across various specialized domains including natural language processing (NLP), computer vision, speech processing, video generation, and code synthesis. However, most existing AI tools are developed in isolation, limiting their capability to address complex, real-world problems that require a combination of modalities. This gap has created a need for an integrated platform that can simultaneously leverage multiple AI models.

This paper introduces a unified AI integration model designed to bring together multiple AI capabilities—text, image, video, audio, and programming code—into a single, interactive platform. The system is particularly tailored for educational and content creation purposes, empowering users such as students, teachers, and digital creators to generate and consume knowledge across modalities without requiring deep technical.

For instance, a user can input a text prompt to generate an explanatory video, visualize the concept with diagrams or illustrations, receive a spoken version of the content for accessibility, and even generate relevant programming code for simulations or technical exercises. Such multimodal synergy enhances the learning experience, promotes creativity, and broadens the usability of AI in everyday tasks.

II. BODY OF THE PAPER

A. Overview of Multimodal AI

Multimodal AI refers to systems that can understand and generate content across different types of data—such as text, images, videos, speech, and code. Each of these modalities contributes uniquely to human expression and learning. A truly intelligent platform must therefore support interaction between these modes to simulate how humans communicate, learn.

This research focuses on integrating five core AI capabilities: Text AI, Image AI, Video AI, Audio AI, and Code AI.

B. System Architecture

The platform is designed using a modular architecture where each AI service operates independently yet communicates through a central orchestrator. This design ensures scalability, interoperability, and flexibility in combining services.

Fig. 2.2 Agile Model Workflow

This diagram illustrates the Agile software development lifecycle, emphasizing an iterative approach through six key phases: Plan, Design, Develop, Test, Deploy, and Review.
Agile promotes continuous improvement, collaboration, and customer feedback to deliver adaptive and high-quality software solutions.

C. Working and Methodology

The Multiple AI Integration Platform is designed to combine the capabilities of diverse artificial intelligence models into a single, interactive interface. The system is modular and service-oriented, enabling seamless interaction with AI agents tailored for specific tasks such as text generation, image analysis/generation, audio processing, video summarization, and code generation.

1. System Architecture Overview

The platform uses a microservices-based architecture, where each AI component is encapsulated in an independent service. These services interact through a central orchestrator or middleware using REST APIs or message queues

2. Component-wise Integration

a. Text AI Module

- Uses LLMs (like GPT or BERT) for tasks such as content creation, summarization, translation, Q&A, and chatbot interactions.
- Receives input from the UI, processes it via a language model, and returns human-like responses.

b. Image AI Module

- Supports both image classification and generation.
- For classification, it uses CNNs trained on labeled datasets.
- For generation, it integrates models like Stable Diffusion or DALL·E to generate images based on textual prompts.

c. Video AI Module

- Performs video summarization, object detection, scene segmentation, and caption generation.
- Leverages transformer-based video models or pretrained action recognition networks.
- Input video is parsed frame by frame, processed, and relevant insights are rendered.

d. Audio AI Module

- Converts speech to text (using ASR like Whisper) and text to speech (using TTS like Tacotron or WaveNet).
- Also supports sentiment analysis from voice tone and speaker diarization.

e. Code AI Module

- Provides code generation, explanation, and debugging.
- Integrates LLMs fine-tuned on programming languages like Python, Java, C++, etc.
- Supports syntax validation and runtime simulation in a sandboxed environment.

3. Unified User Interface

The platform's frontend presents a dashboard-style interface where users can:

- Choose the AI service (Text, Image, Audio, Video, Code).
- Input data or prompt.
- View output in real-time or download results.

It also includes:

- File upload support
- Drag-and-drop input
- API access for external applications

4. Workflow Pipeline

1. Input Handling – User submits input through the web interface.
2. Service Routing – Middleware routes input to the appropriate AI module.
3. Processing – The selected AI model performs its task.
4. Response Aggregation – Results are formatted and returned to the UI.
5. User Feedback – Option for users to rate the output, aiding continuous improvement.

5. Deployment & Tools

- Backend: Python (FastAPI/Flask), Dockerized microservices.
- Frontend: React/Angular.
- ML Libraries: PyTorch, TensorFlow, Hugging Face Transformers, OpenCV.
- Orchestration: Kubernetes (for scaling), Nginx (as reverse proxy), Redis (for session caching).
- Cloud Deployment: AWS/GCP for model hosting and scaling compute resources.

6. Security and Privacy

- Input data is handled securely and stored temporarily.
- User authentication and API rate limiting are implemented to prevent abuse.
- End-to-end encryption is enabled for data in transit.

D. Workflow and Modal Interactions

The system supports various interactions between AI models. For example, text-to-image generation, voice-to-code synthesis, and image-to-text captioning enable comprehensive, multimodal learning and content production.

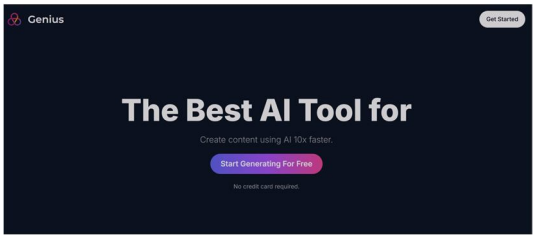


Fig. 2.3.1 Genius AI- Home Page

Welcome to Genius – Your all-in-one AI companion to generate content 10x faster.

Start creating text, images, videos, and more

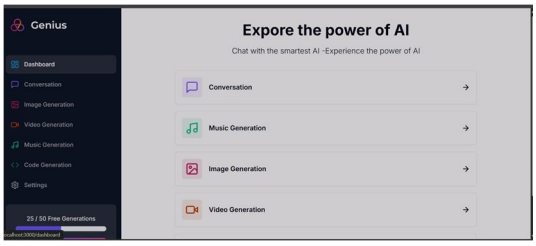


Fig. 2.3.2 Dashboard

This is the main dashboard of Genius AI, showcasing its core features—conversation, music, image, and video generation. Users can interact with powerful AI modules for multimodal content creation, all from a unified interface.

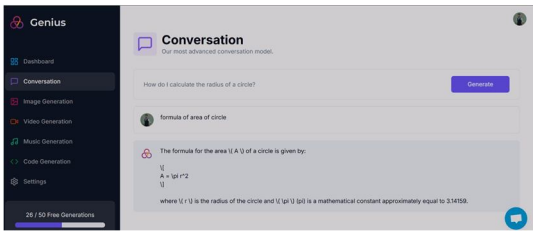


Fig. 2.3.3 Text Conversation

This page highlights Genius AI's conversation model, designed to answer user queries in natural language. Here, the user interacts with the AI to solve a mathematical problem, showcasing its capability in real-time Q&A and educational support.

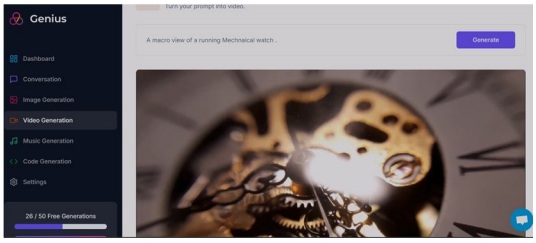


Fig. 2.3.4 Video Generation

This page showcases the video generation capability of Genius AI, where users can create realistic videos from text prompts. In this example, the AI has generated a detailed macro view of a mechanical watch in motion, demonstrating its visual storytelling potential.

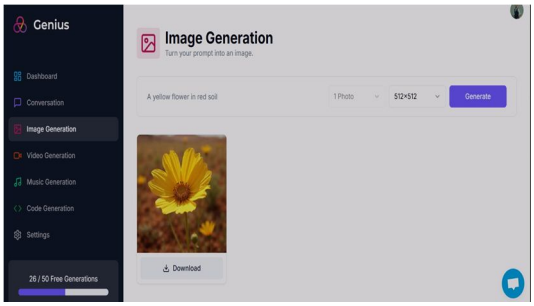


Fig. 2.3.5 Image Generation

This page features Genius AI's image generation capability, allowing users to turn text prompts into high-quality visuals. In this instance, the prompt 'a yellow flower in red soil' has been transformed into a vivid and accurate AI-generated image.

E. Use Case: Education and Content Creation

In educational settings, this system allows teachers to input a topic and receive outputs such as written summaries, narrated explanations, generated infographics, short videos, and executable code, all from a single prompt.

F. Technical Challenges

Challenges include model alignment, latency, data handling, and semantic consistency across modalities. Addressing these requires efficient architecture and robust data exchange protocols.

G. Future Scope

The platform can evolve to serve healthcare, personalized learning, assistive technology, and enterprise use cases, expanding its utility far beyond the initial education and content creation focus.

III. CONCLUSION

The convergence of multiple artificial intelligence domains into a single integrated platform marks a transformative shift in how users interact with technology. This paper has presented a unified AI model that combines the capabilities of text, image, audio, video, and code generation into a cohesive system. By leveraging the strengths of each modality, the platform facilitates seamless cross-modal interactions, enabling users to create, learn, and communicate more effectively.

Our proposed system demonstrates how multimodal AI can significantly enhance educational experiences and content production workflows. Through intelligent orchestration of specialized models, users can input a simple prompt and receive diverse, meaningful outputs. While the system introduces several technical and design challenges, our modular architecture offers a scalable and adaptable solution. In conclusion, this integration has the potential to democratize advanced AI capabilities and pave the way for more accessible, creative, and human-centric AI systems.

IV. ACKNOWLEDGMENT

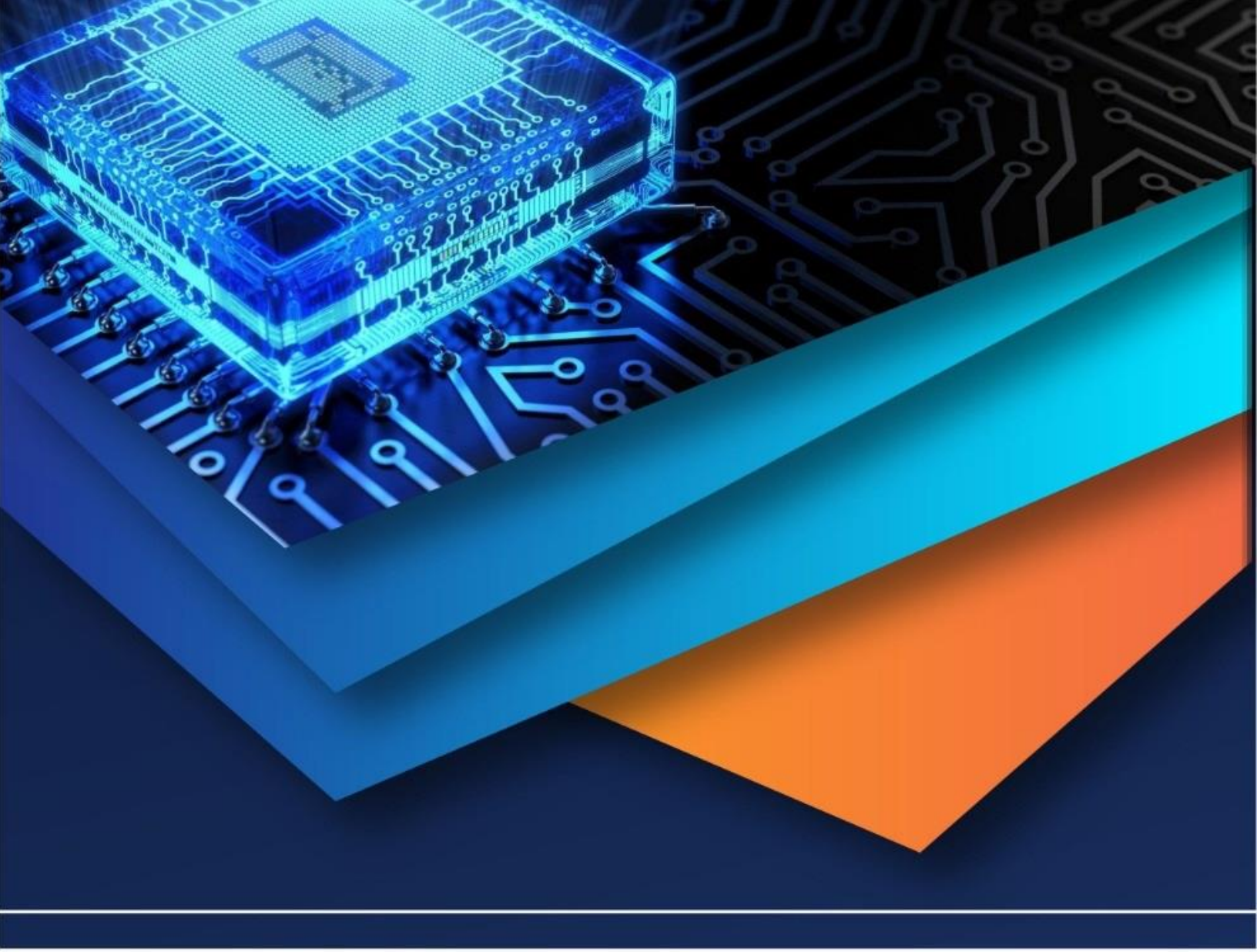
The authors would like to express their sincere gratitude to the faculty and research staff at [Your Institution/University Name] for their continuous support and valuable feedback throughout the development of this project. Special thanks to the AI research community and open-source contributors whose work on large language models, computer vision frameworks, and multimodal integration technologies laid the foundation for this research. We also acknowledge the support of tools and APIs that facilitated the development and testing of the platform.

REFERENCES

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
[2] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI.
[3] Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-shot text-to-image generation. International Conference on Machine Learning (ICML).
[4] Kim, Y., Jernite, Y., Sontag, D., & Rush, A. M. (2016). Character-aware neural language models. AAAI Conference on Artificial Intelligence.
[5] Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7794-7803.
[6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL-HLT, 4171–4186.
[7] OpenAI. (2023). ChatGPT and GPT-4 Technical Report. OpenAI Technical Reports. Retrieved from https://openai.com/research/gpt-4
[8] Google Research. (2023). Gemini: A multimodal AI model. Retrieved from https://deepmind.google
[9] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28.
[10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

BIBLIOGRAPHIES

[1] **Mr. Sahil Meshram** Currently pursuing Bachelor of Technology in Computer Science Engineering from Government Engineering College, Bilaspur Chhattisgarh. Area of interest is Web Development and coding in DSA.
[2] **Mr. Shivam Sahu** Currently pursuing Bachelor of Technology in Computer Science Engineering from Government Engineering College, Bilaspur Chhattisgarh. Area of interest is Web Development and coding in DSA.
[3] **Mr. Shivam Kumar Gupta** Currently pursuing Bachelor of Technology in Computer Science Engineering from Government Engineering College, Bilaspur Chhattisgarh. Area of interest is Web Development and coding in DSA.
[4] **Mr. Yatharth Sonteke** Currently pursuing Bachelor of Technology in Computer Science Engineering from Government Engineering College, Bilaspur Chhattisgarh. Area of interest is Web Development and coding in DSA.
[5] **Mr. Nawaz Mehmood** Currently pursuing Bachelor of Technology in Computer Science Engineering from Government Engineering College, Bilaspur Chhattisgarh. Area of interest is Web Development and coding in DSA.
[6] **Ms. SAVITA SAHU** Currently working as a Assistant Professor in Computer Science Engineering in Government Engineering College, Bilaspur Chhattisgarh. Area of interest is Data Mining, Artificial Intelligence, Machine Learning.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY