IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 2, March 2025

Visual Genesis AI

Deveshree Wankhede[1] Leena Bonde[2], Sumeet Borole[3], Anish Lad[4]

Faculty, Department of Computer Engineering [1]

Student, Department of Computer Engineering [2,3,4]

Shivajirao S Jondhale College of Engineering, Dombivli (E),Thane, Maharashtra, India

Abstract: *VISUAL GENESIS AI is an advanced multimedia generation platform that utilizes artificial intelligence (AI) to create images, videos, and music in a user-friendly and efficient manner. The system integrates cutting-edge AI technologies, including deep learning models such as Generative Adversarial Networks , Transformer-based models, and Long Short-Term Memory networks, to automate the creative process. Designed with a Node.js backend for real-time performance and scalability, VISUAL GENESIS AI simplifies the creation of high-quality digital media for users across diverse industries, from entertainment and advertising to education and content creation.*

The project addresses the increasing demand for AI-assisted creative tools that can meet the growing expectations of digital content consumers. By providing an integrated platform that supports multiple media types, VISUAL GENESIS AI distinguishes itself from existing single media tools, offering an all-in-one solution for creators who need to generate a variety of content types. Users can generate images from textual descriptions, synthesize original video clips from images or video prompts, and compose music tracks based on a range of parameters such as mood, genre, and tempo.

This platform is especially useful for those without specialized knowledge in media production or AI, as its intuitive user interface and underlying automation reduce the complexity traditionally associated with these tasks. The system also leverages the capabilities of Node.js for seamless performance and scalability, ensuring that even large-scale projects can be executed efficiently.

Keywords: Generative Adversarial Networks(GANs), Variational Autoencoder (VAE), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Artificial Intelligence (AI), Long Short Term Memory (LSTM)

I. INTRODUCTION

1.1 Overview

In the rapidly evolving landscape of artificial intelligence, the capacity to generate and manipulate multimedia content has reached unprecedented heights. "VISUAL GENESIS AI" is a pioneering project that harnesses the power of AI to revolutionize the creation of images, videos, and music. Utilizing advanced algorithms and Node.js, this project aims to bridge the gap between human creativity and machine intelligence, offering a robust platform for seamless multimedia generation.

As technology continues to reshape the boundaries of artistic expression, "VISUAL GENESIS AI" seeks to explore the potential of AI-driven creativity, providing users with innovative tools to generate high-quality content across various media. This introduction will delve into the core objectives of the project, the underlying technologies, and the potential impact on creative industries and beyond.

In a world where technology and creativity collide, "VISUAL GENESIS AI" is here to make art and media creation easier and more exciting than ever. Imagine having a tool that not only understands your creative ideas but also brings them to life with the help of advanced AI. This project combines cutting-edge AI with Node.js to let you generate stunning images, captivating videos, and unique music. Whether you're an artist looking for inspiration or just curious about what AI can create, "VISUAL GENESIS AI" offers a glimpse into the future of creative technology.

DOI: 10.48175/IJARSCT-23730

1.2 Problem Statement

Existing AI models for image generation face several challenges. One major issue is **limited creativity**, as they often struggle to produce truly unique and original images, sometimes generating repetitive or predictable results. Another challenge is **controllability issues**, where users may find it difficult to direct the AI model to achieve their desired outcomes. Additionally, **ethical concerns** arise when AI-generated visuals are misused for malicious purposes, such as creating harmful content. Moreover, some AI models suffer from **limited resolution**, leading to blurry outputs, while others exhibit a **lack of detail**, failing to capture fine elements and reducing realism. Lastly, **computational requirements** pose a significant hurdle, as training and running AI models for image and video generation demand powerful hardware, making them resource-intensive.

1.3 Objectives

- Use a combination of GANs (Generative Adversarial Networks) and Diffusion Models to enhance creativity.
- Introduce style mixing and latent space exploration to produce diverse outputs.
- Provide users with adjustable parameters.
- Enable user-selectable resolutions, allowing them to request higher-quality outputs.
- Use Hybrid AI models.

II. LITERATURE REVIEW

1) Paper Title: "DALL·E: Zero-Shot Text-to-Image Generation"(2021)
Authors: Ramesh.
The capabilities of GPT-like architectures in multimodal tasks are showcased by the model that generated diverse and high-quality images from text. Difficulties in producing fine-grained details and artifacts in some generated images. Transformer-based architecture, trained on large datasets of image-text pairs, enabling it to generate coherent images from text prompts.

2) Paper Title: "VideoGPT: Video Generation with a Generative Pretrained Transformer"(2021)
Authors: Yan,Villegas,Lee.
Developed a model capable of producing diverse and high-quality images from textual descriptions, demonstrating the potential of GPT-like architectures in multimodal applications. The model exhibits impressive zero-shot performance, generating varied images from text prompts with minimal fine-tuning. However, challenges remain in rendering fine-grained details, with occasional artifacts appearing in the outputs. The system employs a transformer-based architecture trained on extensive image-text datasets, enabling it to create coherent and contextually accurate images from textual inputs.

3) Paper Title: "Jukebox: A Generative Model for Music"(2020)
Authors: Dhariwal, P., Jun.
Created a high-fidelity, multi-genre music generation model capable of producing full songs, including vocals and instrumental arrangements. The model excels in generating high-quality music across various genres, incorporating complex compositions with lyrics and harmonies. However, it is highly computationally intensive and requires substantial datasets for training. The architecture leverages a VQ-VAE-based approach combined with autoregressive and hierarchical generative modeling to produce long-form music with intricate structures.

4) Paper Title: "GANs for Video Generation and Prediction"(2020)
Authors: Clark, A., Donahue.
Showcased the application of GANs in video generation and prediction, with a focus on enhancing the temporal consistency of generated sequences. Achieved improved quality and coherence across video frames by effectively capturing spatial and temporal dynamics. However, challenges persist in producing high-resolution videos over extended durations. The model extends traditional image generation techniques to video, leveraging GAN-based architecture with recurrent or convolutional layers to maintain temporal consistency.

5) Paper Title: "StyleGAN2: Analyzing and Improving the Image Quality of StyleGAN”(2020)
Authors: Karras, T., Laine.
Enhanced StyleGAN by addressing common artifacts and boosting the overall realism and quality of generated images. This advancement significantly reduced visual artifacts and improved sharpness and detail in the outputs. Despite these improvements, the model still demands substantial computational resources and large datasets for effective training. The refined architecture incorporates advanced normalization techniques, such as weight demodulation, to enhance image quality and maintain consistency.

6) Paper Title: "StyleGAN: A Style-Based Generator Architecture for GANs”(2019)
Authors: Karras, T., Laine.
Introduced a style-driven approach to image generation, enabling precise control over various elements of the image, such as facial features and backgrounds. This approach allows for detailed manipulation of image attributes through style mixing, enhancing the interpretability of the generated features. However, challenges with training stability remain, occasionally resulting in visual artifacts. The architecture employs a style-based generator, where input styles influence specific image characteristics via intermediate layers.

7) Paper Title: "WaveNet: Generative Model of Raw Audio”(2016)
Authors: Aaron van den.
WaveNet is a deep generative model designed for synthesizing raw audio waveforms, renowned for its ability to produce high-quality, realistic audio. This model has had a significant impact on music and speech generation, achieving state-of-the-art results in audio synthesis with natural and high-fidelity outputs. Its architecture is versatile and can be adapted for a range of audio-related applications. However, the model’s high computational requirements and training complexity present challenges. Additionally, its performance heavily depends on the size and quality of the training dataset. WaveNet utilizes a stack of convolutional layers to sequentially generate audio samples, effectively capturing fine-grained audio details and long-term temporal dependencies.

8) Paper Title: "Deep Video Generation with Recurrent Neural Networks"(2016)
Authors: Xue Bin Peng.
This paper explores the use of recurrent neural networks (RNNs) for video sequence generation, with a focus on ensuring temporal consistency and realistic motion in the generated videos. It presents a framework for creating videos that maintain coherence over time, addressing the challenge of temporal continuity in video generation. Video generation remains computationally intensive and may suffer from limitations in resolution and detail. The approach leverages RNNs to capture temporal dependencies across video frames, combining both spatial and temporal features to produce lifelike video content.

9) Paper Title: "Generative Adversarial Networks for Image Generation"(2014)
Authors: Ian Goodfellow et al.
This pioneering paper introduces Generative Adversarial Networks (GANs), a framework in which two neural networks compete to generate realistic images. GANs have become a cornerstone in the field of image generation. The paper demonstrated GANs' capability to produce high-quality, highly realistic images, and the method has since been widely adopted and expanded across various applications. Despite their success, GANs can be difficult to train and may produce artifacts or lack diversity in the generated outputs. GANs consist of two networks: a generator that creates images and a discriminator that evaluates their authenticity, pushing the generator to enhance its performance.

10) Paper Title:-"Auto-Encoding Variational Bayes"(2014)
Authors:-Kingma
This work introduces the Variational Autoencoder (VAE) framework, which combines variational inference with autoencoders for generative tasks. The VAE enables unsupervised learning and can generate new data—such as images and videos—by sampling from a learned latent space. While it is effective, the generated images are often blurry and

the model struggles with producing high-resolution outputs compared to GANs. The encoder produces a probabilistic distribution (mean and variance) over the latent variables, while the decoder reconstructs data from this latent space. The loss function balances reconstruction error with KL divergence to regularize the latent space.

III. PROPOSED SYSTEM

3.1 Introduction

The proposed system of VISUAL GENESIS AI consists of three core functionalities:

Image Generation: The system leverages a pre-trained neural network model, such as GANs or Diffusion Models, to create high-quality images from textual descriptions or image prompts. The process is handled on the backend using Node.js, ensuring quick processing and a seamless user experience.

Video Generation: For video content, VISUAL GENESIS AI utilizes AI models that can either convert images into video sequences or create entirely new video clips from a series of instructions. Methods like GAN-based video prediction and video generation using recurrent neural networks are implemented to achieve this.

Music Generation: In the music domain, the system offers users tools to create original tracks using AI-powered models, including those based on deep learning architectures like LSTMs and Transformer models. The generated music can be shaped by different parameters such as mood, genre, and rhythm.

Code Generation: In the coding domain, the system provides users with AI-powered tools to generate high-quality code across various programming languages. Leveraging advanced deep learning models, such as Transformer-based architectures, the platform enables users to create functional scripts, automate tasks, and develop applications efficiently.

Conversation: In the conversation domain, the system utilizes AI-driven natural language processing (NLP) models to facilitate dynamic and context-aware interactions. Powered by advanced deep learning architectures like Transformers, the platform enables seamless and intelligent conversations, whether for customer support, creative brainstorming, or general discussions.

Secure & Scalable Database Management (Prisma + PostgreSQL): To store and manage user-generated content, Prisma ORM and PostgreSQL work together to provide a structured, scalable, and efficient database solution. Prisma simplifies database interactions with a type-safe API, making queries easy to write and maintain. PostgreSQL, a powerful relational database, ensures high performance, security, and reliability, even with large-scale AI-generated data. Together, they provide a robust backend for seamless data storage and retrieval.

Seamless Payment Integration (Stripe): For monetization, Stripe is integrated to handle transactions securely. It enables subscription-based access, one-time payments, and automated invoicing, ensuring users can effortlessly purchase AI-generated assets while benefiting from top-tier security and fraud protection mechanisms. Stripe's API allows smooth financial transactions while ensuring compliance with global payment regulations.

DOI: 10.48175/IJARSCT-23730

3.2 System Architecture

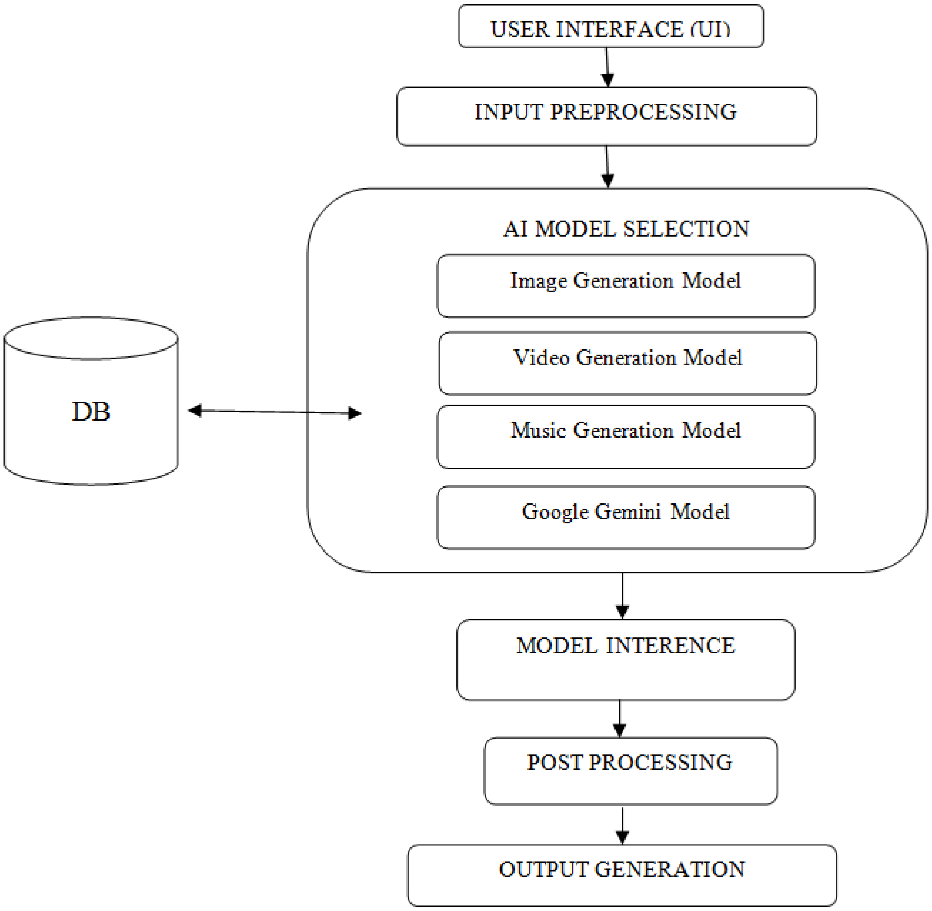


Fig 3.2.1 System Architecture

3.3 Each Model with Alg

In the rapidly evolving landscape of AI, Visual Genesis stands out as an all-in-one platform that brings together text, image, video, music, and code generation—all within a unified system. By leveraging cutting-edge AI models and seamless integrations, this project enables users to interact with an intelligent assistant, generate high-quality multimedia content, and manage data efficiently. Whether it’s creating professional visuals, generating music, writing code, or handling transactions, Visual Genesis offers a streamlined and intuitive experience.

Working

1. Conversational AI & Code Generation (Google Gemini)

At the core of intelligent interactions, Google Gemini powers the system’s conversational and code-generation capabilities. Built on a Transformer-based deep learning architecture, Gemini employs Mixture of Experts (MoE) techniques to enhance efficiency and accuracy. It understands user queries, generates coherent responses, and even writes optimized code based on the latest LLM (Large Language Model) advancements.

2. AI-Powered Image Generation (Together AI + Flux Dev Schnell Model)

To create stunning visuals from text prompts, Visual Genesis integrates Together AI, which hosts the Flux Dev Schnell model. This model is built on a Diffusion Transformer (DiT) architecture and trained using CLIP-guided latent diffusion. By progressively refining noise into structured images, it produces high-resolution, photorealistic results that align closely with the user’s input.

DOI: 10.48175/IJARSCT-23730

3. Advanced Video Generation (Replicate AI + Lightricks/LTX-Video)

For video content, Lightricks/LTX-Video, hosted on Replicate AI, enables text-to-video generation. It utilizes a combination of latent diffusion models (LDMs) and spatiotemporal attention mechanisms to maintain frame consistency and motion accuracy. This ensures the generated videos are not only visually appealing but also maintain natural transitions and temporal coherence.

4. AI-Generated Music Composition (Replicate AI + Riffusion Model)

Music generation is powered by Riffusion, a model based on Stable Diffusion but trained on spectrogram representations of audio. It transforms text-based prompts into waveform-like spectrograms, which are then converted back into music using a synthesis pipeline. This allows users to generate diverse, genre-specific music with minimal input.

5. Secure & Scalable Database Management (Prisma + PostgreSQL)

To store and manage user-generated content, Prisma ORM and PostgreSQL work together to provide a structured, scalable, and efficient database solution. Prisma simplifies database interactions with a type-safe API, ensuring smooth CRUD operations, while PostgreSQL ensures reliability and high performance, even under heavy workloads.

6. Seamless Payment Integration (Stripe)

For monetization, Stripe is integrated to handle transactions securely. It enables subscription-based access, one-time payments, and automated invoicing, ensuring users can effortlessly purchase AI-generated assets while benefiting from top-tier security and fraud protection mechanisms.

IV. METHODOLOGY

User Input: The initial stage of the system, where the user inputs parameters or selects the type of content to generate.

Preprocessing: Processes the user's input by normalizing or converting it into a format that is compatible with the AI models (e.g., resizing images or encoding audio).

AI Models:

Image Generation Model– For generating images based on user input.

Video Generation Model– For generating video sequences.

Music Generation Model– For generating music.

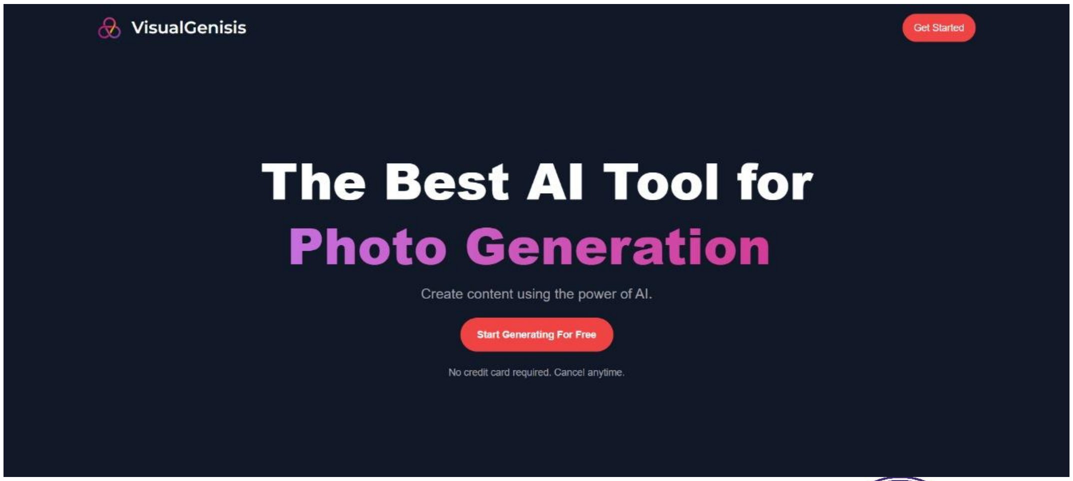
Conversation Model- For seamless and intelligent conversation.

Code Generation Model- For generating codes in different languages.

Post-processing: Refines or optimizes the raw outputs.

Output: The generated content (whether image, video, or music) is presented to the user or stored for later use.

V. RESULTS



VisualGenisis

Get Started

The Best AI Tool for
Photo Generation

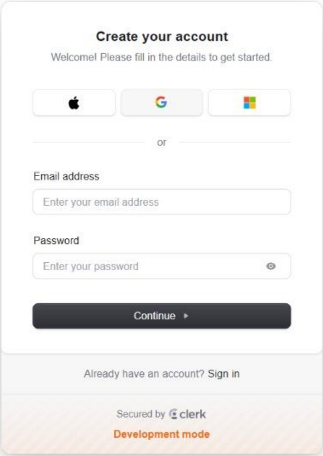
Create content using the power of AI.

Start Generating For Free

No credit card required. Cancel anytime.

5.1 Landing Page

Here is the image of the landing page interface for the VISUAL GENESIS AI platform, featuring a sleek and user-friendly layout for generating images, videos, and music.



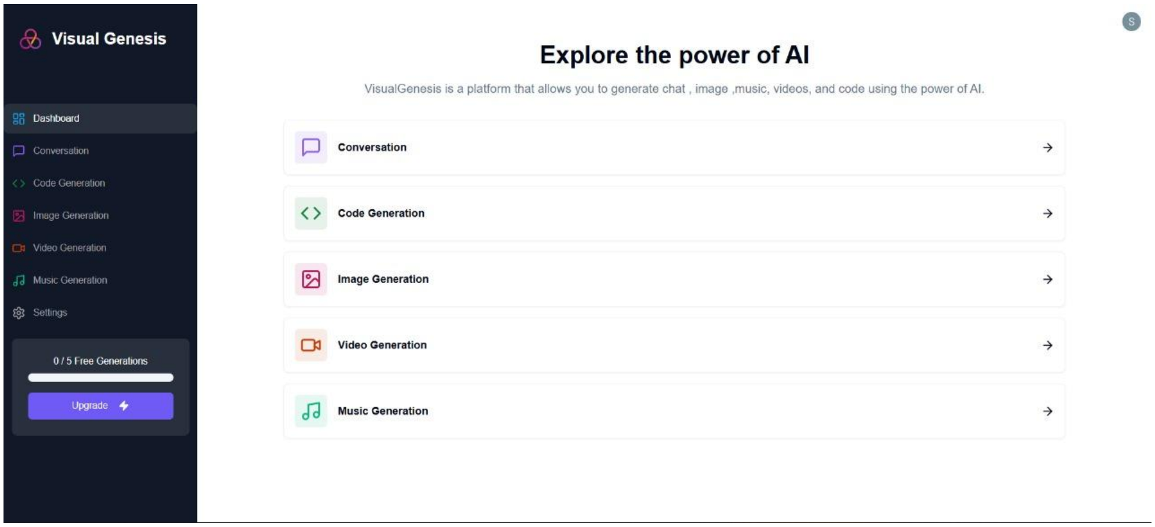
Create your account

Welcome! Please fill in the details to get started.

Already have an account? Sign in

5.2 Signup Page

Here is the image of the "Create Account" page interface for the VISUAL GENESIS AI platform, showcasing a sleek and intuitive design. The layout focuses on simplicity and user convenience.



Explore the power of AI

VisualGenesis is a platform that allows you to generate chat , image ,music, videos, and code using the power of AI.

5.3 Dashboard

Your Dashboard module provides users with a centralized interface to access and manage all AI models, enabling seamless interaction with image, video, and music generation tools.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 2, March 2025

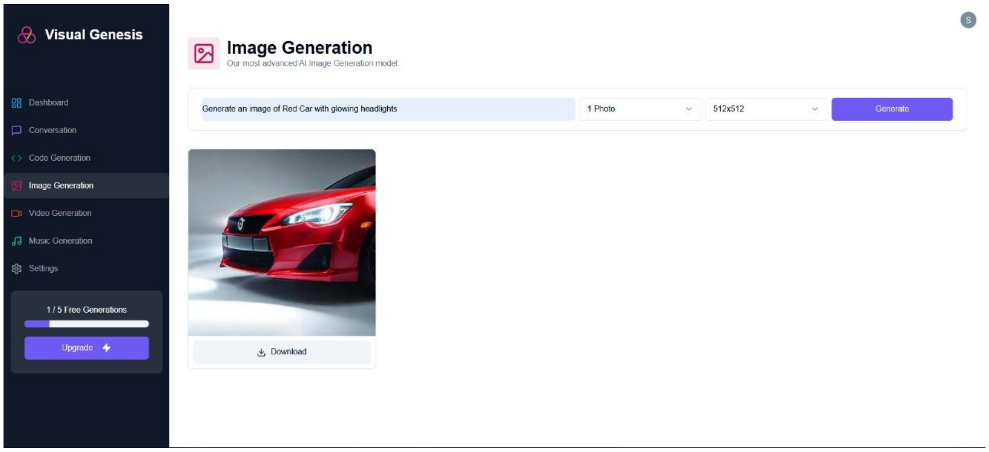


Image Generation
Our most advanced AI Image Generation model.

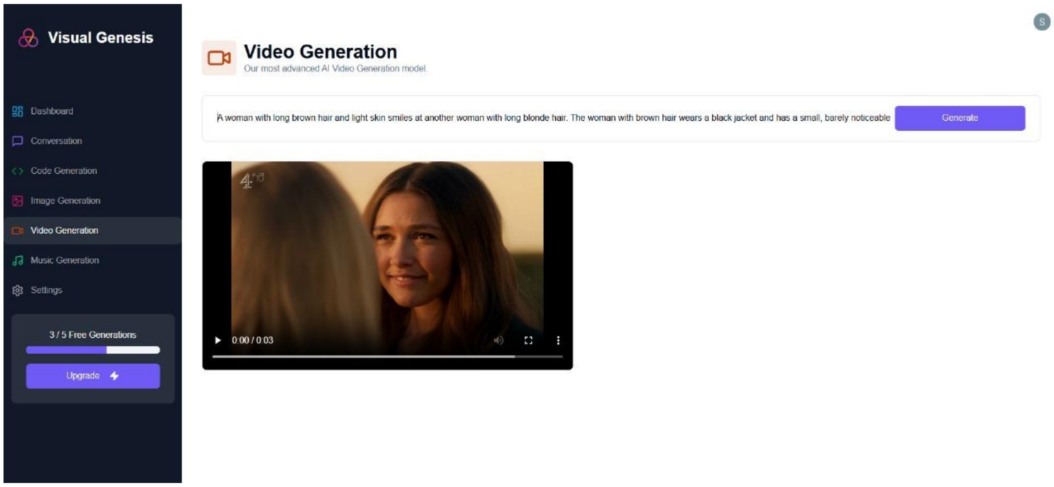
Generate an image of Red Car with glowing headlights

1 Photo

512x512

5.4 Image Generation Module

The image generation module enables users to input text or prompts to quickly generate high-quality AI images. It offers customization options for style, resolution, and effects, providing creative results with ease.

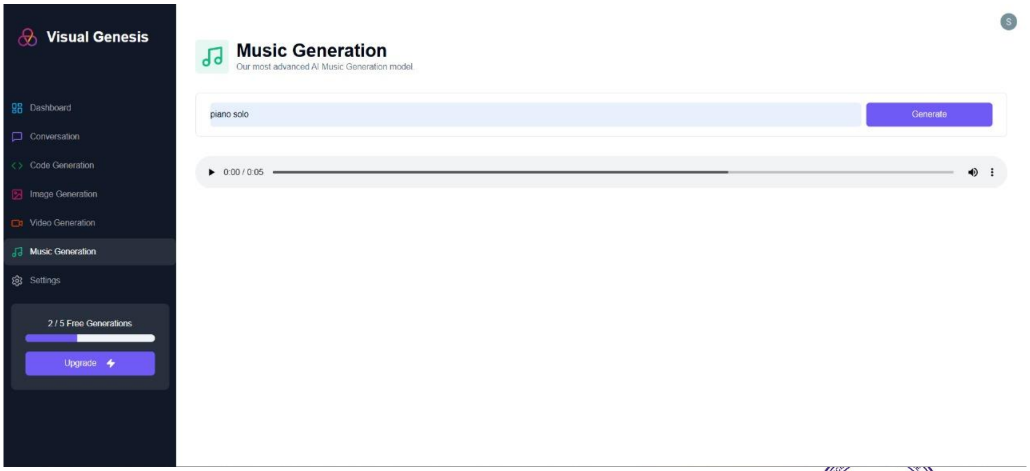


Video Generation
Our most advanced AI Video Generation model.

A woman with long brown hair and light skin smiles at another woman with long blonde hair. The woman with brown hair wears a black jacket and has a small, barely noticeable

5.5 Video Generation Model

The video generation module allows users to create AI-powered videos from text-based input, offering customization options for style, resolution, and effects to produce high-quality, dynamic content.



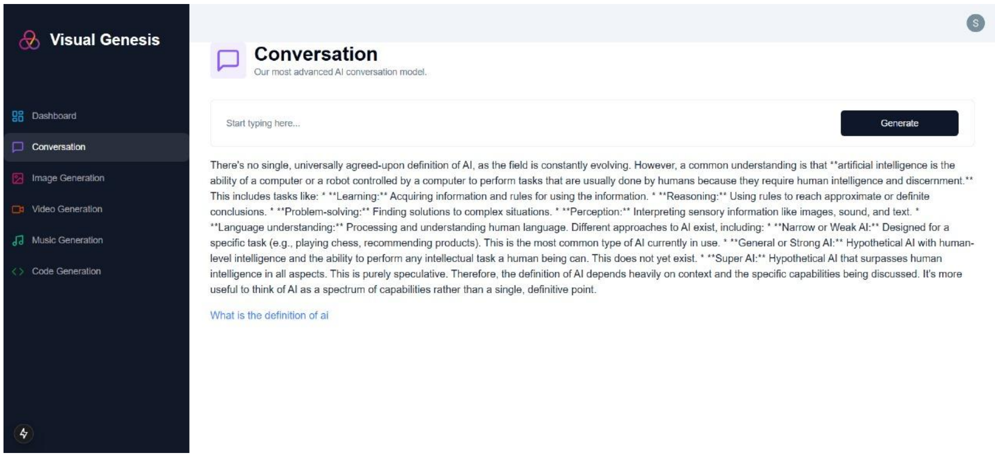
Music Generation
Our most advanced AI Music Generation model.

piano solo

5.6 Music Generation Model

DOI: 10.48175/IJARSCT-23730

The music generation module allows users to create AI-powered music from text-based input, offering customization options for genre, tempo, and mood to produce high-quality, unique compositions

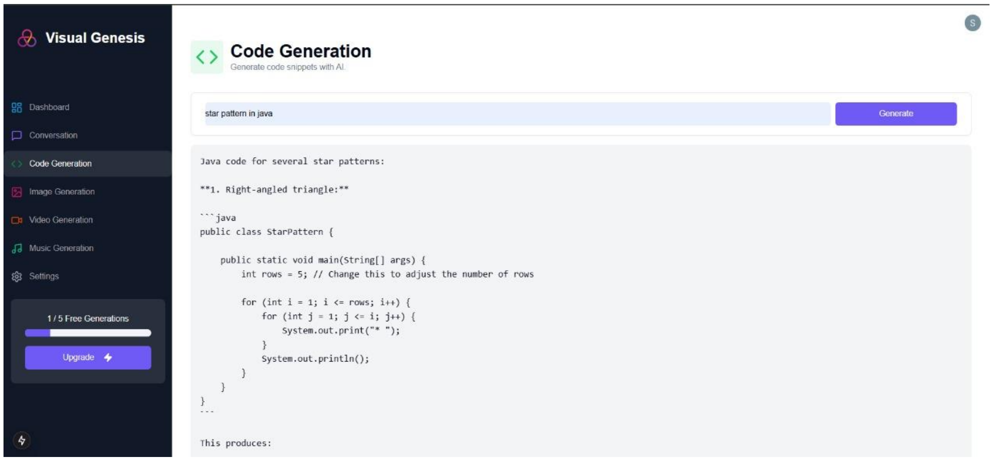


There's no single, universally agreed-upon definition of AI, as the field is constantly evolving. However, a common understanding is that **artificial intelligence is the ability of a computer or a robot controlled by a computer to perform tasks that are usually done by humans because they require human intelligence and discernment.** This includes tasks like: * **Learning:** Acquiring information and rules for using the information. * **Reasoning:** Using rules to reach approximate or definite conclusions. * **Problem-solving:** Finding solutions to complex situations. * **Perception:** Interpreting sensory information like images, sound, and text. * **Language understanding:** Processing and understanding human language. Different approaches to AI exist, including: * **Narrow or Weak AI:** Designed for a specific task (e.g., playing chess, recommending products). This is the most common type of AI currently in use. * **General or Strong AI:** Hypothetical AI with human-level intelligence and the ability to perform any intellectual task a human being can. This does not yet exist. * **Super AI:** Hypothetical AI that surpasses human intelligence in all aspects. This is purely speculative. Therefore, the definition of AI depends heavily on context and the specific capabilities being discussed. It's more useful to think of AI as a spectrum of capabilities rather than a single, definitive point.

What is the definition of ai

5.7 Conversation Page

The Conversation Module enables users to text or prompt to getting solutions.



star pattern in java

~~~~
Java code for several star patterns:

**1. Right-angled triangle:**

```java
public class StarPattern {

    public static void main(String[] args) {
        int rows = 5; // Change this to adjust the number of rows

        for (int i = 1; i <= rows; i++) {
            for (int j = 1; j <= i; j++) {
                System.out.print("* ");
            }
            System.out.println();
        }
    }
}
```

This produces:
~~~~

5.8 Code Generation Model

The code generation module allows users to create AI-powered code from text-based input, offering customization options for programming languages, frameworks, and complexity to generate efficient and high-quality code.

VI. CONCLUSION

VISUAL GENESIS AI seeks to revolutionize digital content creation by offering an integrated platform for AI-driven image, video, and music generation. It connects creativity with technology, making the creation of high-quality media accessible to everyone. With ongoing advancements in AI models and opportunities for growth, VISUAL GENESIS AL has the potential to become an invaluable tool for both professional creators and enthusiasts across various industries. Future updates may include expanding the types of supported media and enhancing the model's creative capabilities

The future scope of **VISUAL GENESIS AI** is vast, with numerous opportunities for growth and innovation. Advancements in AI models can enhance creativity by integrating multimodal AI, allowing seamless generation of images, videos, and music. Features like AI-assisted storyboarding, real-time editing tools, and conditional generation will provide users with greater control and customization. To improve realism, the platform can implement super-resolution techniques, motion interpolation for videos, and detailed enhancement models. Ethical AI development will also be a priority, ensuring fairness, watermarking for transparency, and content moderation to prevent misuse.

DOI: 10.48175/IJARSCT-23730

ACKNOWLEDGMENT

We hereby declare that the work presented in this Mini project report entitled:**"** in partial fulfilment for the Degree of **"Bachelor of Engineering"** in **Computer Engineering**.

Many people have contributed to our achievements during the project and we take this opportunity to thank each one of them. First, we would like to thank our Principal **Dr. Pramod R.Rodge**, Head Of Department **Dr. Uttara Gogate** and our project coordinator **Prof. Reena Deshmukh** who has extended all her valuable guidance, help and constant encouragement through the various difficult stages involved in the development of the project.

Also, we would like to specially thank **Prof. Deveshree Wankhede** who took her precious time out from her busy schedule for our major project and helped us a ton by being the project guide. Without their willing disposition, spirit of accommodation, frankness, timely clarification and above all faith in us, this project could not have been completed in due time.

REFERENCES

[1]. Ramesh . “DALL•E: Zero-Shot Textto-Image Generation”, arXiv preprint , 2021.

[2]. Yan, Villegas, Lee, "VideoGPT: Video Generation with a Generative Transformer", IEEE Conference on CVPR, 2021.

[3]. Dhariwal, P., Jun "Jukebox: A Generative Model for Music",IEEE Conference on CVPR,arXiv preprint, 2021.

[4]. Clark, A., Donahue,"GANs for Video Generation and Prediction",International Conference on Learning Representations (ICLR),2020.

[5]. Karras, T., Laine,"StyleGAN2: Analyzing and Improving the Image Quality of StyleGAN", IEEE International Conference on Computer Vision (ICCV)”,2020.

[6]. Karras, T., Laine,"StyleGAN: A Style-Based Generator Architecture for GANs", 2020

[7]. Aaron van den, "WaveNet: A Generative Model for Raw Audio.”, IEEE Conference on CVPR , 2016.

[8]. XueBin Peng, "Deep Video Generation with Recurrent Neural Networks.", IEEE Conference on CVPR, 2016.

[9]. Good fellow Ian ,"Generative Adversarial Nets." Advances Neural Information Processing Systems, 2014.

[10]. Kingma, D.P., and Welling, M. "Auto-Encoding Variational Bayes." ICLR , 2014

DOI: 10.48175/IJARSCT-23730