International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 11 | Nov 2024 www.irjet.net p-ISSN: 2395-0072

JINNY AI: SURVEY ON AI FOR IMAGE, VIDEO AND AUDIO GENERATION

Devashish Potnis[1], Prathmesh Chavan[2], Abhishek More[3], Sudesh Patil[4], Prof. Mrs. Sujata Sonawane[5]

[1]*Assistant Professor, Dept. of Artificial Intelligence & Machine Learning Engineering, PES's Modern College of Engineering, Pune, Maharashtra, India*
[2,3,4,5]*Student, Dept. of Artificial Intelligence & Machine Learning Engineering, PES's Modern College of Engineering, Pune, Maharashtra, India*

Abstract - *This paper introduces and develops an AI platform to generate high-quality images, videos, and music based on natural language prompts. This platform is simply using the strong ability of state-of-the-art AI models and technologies, focusing on understanding and semantics of user input for relevant outputs. The paper outlines architecture, generative models, user interface design, and applications in different industries and underlines the importance of creativity and efficiency in multimedia content. This work is aimed at providing in-depth analysis of the relevant theory concepts, work, and future directions, that way availing capable insights into the advancement of capabilities in AI-driven content generation.*

Key Words: Artificial Intelligence, Image Generation, Audio Generation, Video Generation, Open AI, MERN Stack, Next.js, React, Tailwind, Prisma, Cleark Authentication.

1. INTRODUCTION

Welcome to the future of software development, where innovation meets artificial intelligence! Our cutting-edge platform combines the power of advanced AI technologies with the creativity and expertise of developers to revolutionize the way applications are built, optimized, and deployed., as our AI platform empowers you to create intelligent, adaptive, and efficient software solutions like never before.

1.1 Context & Motivation

The motivation for this project is to provide users with a powerful tool for creative expression and content generation. By enabling the creation of diverse multimedia content from simple text prompts, the platform aims to democratize access to high-quality content creation tools to streamline the creative process

The aim of this platform is to lessen the harsh separation of technology and creativity by allowing users to create content easier and express their ideas in their own words. There is no need to have outstanding technical skills as the platform allows users to implement their thoughts in real life and enables artists, marketers, teachers, and developers to further enrich their work. In addition, the platform is designed to improve workflow and decrease the resources invested in generating multimedia content, enhancing creativity and enabling more time and attention to be focused on the design and artistic composition of the projects than on their realization.

In short, it's the kind of future when technical constraints no longer become hurdles to creativity. It is that kind of future where AI becomes a co-creator so that users may natively and seamlessly enable their vision. In removing these barriers toward entering multimedia production, a new wave of innovation will thrive behind this platform to help individuals and organizations realize more creative possibilities.

1.2 Problem Definition

It addresses the challenge of creating professional quality multimedia content, including images, video, and audio, all without requiring user expertise or expensive tools. Most existing AI solutions typically handle only one type of media and can't really tell what the user's intent is behind a specific input, leading to inconsistent or even irrelevant outputs. The project's idea is to create an AI content platform that will be capable of creating all kinds of media from simple text prompts such that the produced output will be contextually correct and highly technical in quality. This will allow professional-grade content generation to be accessible to everyone who can simplify the creative process for creation to a great extent.

2. Objectives

- AI platform for high-quality pictures and videos, along with music, from simple text prompts.
- This will ensure correct understanding of the context, and its outputs will be semantically relevant as well as technically precise for all types of media.
- It has several features, such as Clerk authentication to ensure safety access.
- Also, it provides real-time performance for scalable content generation.
- It will ensure a safe and fluid user experience.
- It simplifies the creative process for any user.

3. Review of existing research

The following is Some Important Review existing AI models and technologies used for generating images, videos, and music:-

Tittle	Method	Remark	Technique	Future Scope
AI Based SAAS Project	NextAI is an AI-powered SAAS project that offers a comprehensive suite of creative tools including image generation, video generation, music generation, and conversation AI	Highlights the transformative potential of AI in creative fields, while emphasizing user-centered design and responsible AI.	SAAS Application	Demonstrated the potential impact of NextAI on content creation workflows, emphasizing user control and ethical AI practices.
Applications and Advances of Artificial Intelligence in Music Generation:A Review	Developed a comprehensive summary framework categorizing different technological approaches (symbolic, audio, hybrid). Conducted literature surveys and	Serves as a comprehensive reference for researchers and practitioners in in AI music generation.	Text-to-audio	Proposed future research directions for standardizing evaluation techniques and enhancing application adoption.
A Survey of AI Music Generation Tools and Models	Paper described the music generation technology and its working.	Offers a comprehensive overview of AI music generation, helping users make informed tool choices based on their needs.	Text-to-audio	Further exploration of the evolution of AI music tools, addressing challenges in creativity and user interaction.
Parallel Dense Video Caption Generation with Multi-Modal Features	Introduced a deformable Transformer framework, with an information transfer station with deformable attention for caption generation.	Provides a novel approach to dense video captioning, addressing key issues in event proposal and caption generation.	Text-to-video	Investigate further enhancements in summarization accuracy and explore integration with other media types (e.g., images).
Text-to-Image Diffusion Models	Proposed a low-cost method for video generation that enriches latent codes with motion dynamics and uses cross-frame attention for context preservation.	Provides a comprehensive summary of text-driven image generation methods	Text-to-image	Explore further applications beyond text-to-video synthesis, such as conditional video generation and video editing techniques.

Table -1: Review of Existing Research

Researchers' contribution in AI generates contend:

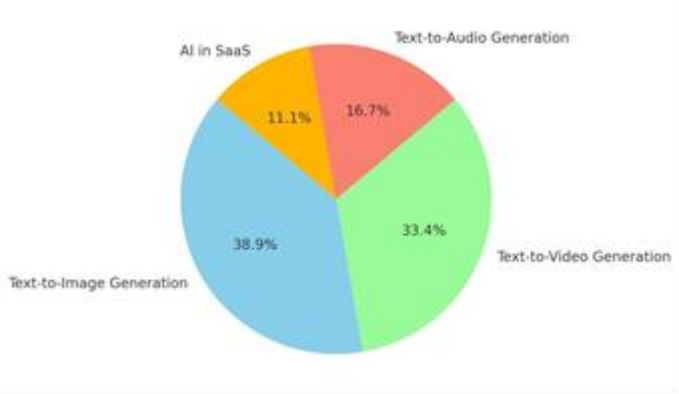

Chart -1: Researchers' contribution in AI generates contend

4. Challenges and Limitations

- Generation Quality and Consistency: It's difficult to guarantee the quality and consistency of images, videos, and music that are generated in response to different types of prompts. Open AI models are very strong, but every once in a while, they'll generate something that doesn't match what the user wants, as far as content or aesthetic style goes.
- Scalability: The processing of high-quality media, including images, videos, and music, may quickly become resource-intensive if done in real-time. Scaling these operations efficiently while holding response times low is a challenge, especially when many users need to interact simultaneously.
- Monetization and Integration: Adding a payment feature might introduce some issues surrounding international payment compliance, currency conversion and transaction security, especially for premium content or service subscriptions that use Stripe.
- Technical Limitations of Current AI Models: The current AI models cannot generate high-fidelity, long-duration videos or complex musical compositions with intricate layers of instruments or vocals. Those would be limitations to overcome toward the desired level of output.
- Data Handling and Security: Data handling in a secure manner- specifically user inputs and generated content-may pose a serious challenge. Clerk authentication might help with many security concerns but further encryption, proper storage of data, and concern for GDPR compliance will be needed.

5. Future Directions

- Custom Model Fine-tuning: Fine-tuning OpenAI models or training proprietary models on user data can help to increase output accuracy, which would in turn bring media generated for these users closer to what users would expect within their field of artistry or technology.
- Advanced Content Customization: Further personalization interfaces for users to customize specific aspects of their output-that is, visual style, for example, or genre of music, or length of video-would make the output more flexible and suitable for creative professionals
- Improving Real-Time Generation: When technology improves, it might better optimize real-time media generation using even more efficient architectures or with the help of edge computing to reduce the computation load on the platform and scale it out and then make it scalable and responsive for a much larger user base

6. CONCLUSIONS

Developing an AI platform that generates images, videos, and music from text prompts, the system needs to understand user input and produce high-quality, relevant content. It should use advanced AI to accurately interpret prompts and create media. The platform must be scalable to handle varying demands while maintaining performance. User data privacy is crucial for security. This will ensure a creative, reliable, and secure experience for users.

REFERENCES

[1] Rashi Malviya, Sakshi Pachlaniya, "AI Based SAAS Project", International Journal of Research Publication and Reviews (IJRPR), vol. 2024

[2] Yanxu Chen, Linshu Huang, Tian Gou, "Applications and Advances of Artificial Intelligence in Music Generation: A Review," arXiv, 2024.

[3] Yueyue Zhu, Jared Baca, "A Survey of AI Music Generation Tools and Models," arXiv, 2023. Santosh Kumar Satapathy, Drashti Parmar, "Video Generation by Summarizing the Generated Transcript," IEEE, 2023.

[4] Hao Sheng, Wei Ke, Ka-Hou Chan, Xuefei Huang, "Parallel Dense Video Caption Generation with Multi-Modal Features," MDPI, 2023

[5] Levon Khachatryan, Andranik Movsisyan, Vahram Tadevosyan, "Text-to-Image Diffusion Models," IEEE, 2023.

[6] Parth Gandhi, Md Dilshad, Vishal Kumar, P. Srilega, "Creato all-in-one SaaS platform with AI-powered tools," International Journal of Advanced Research in Innovative Ideas and Technologies

[7] Hyeonjin Lee, Ubaid Ullah, Jeong-Sik Lee, Bomi Jeong, Hyun-Chul Choi, "A Brief Survey of Text-Driven Image Generation and Manipulation," IEEE, 2021

[8] Olugbenga A. Adenuga, Ray M. Kekwaletswe, "A Systematic Literature Review to Uncover SaaS Adoption Issues by SMEs," ResearchGate, 2020.

[9] Aditi Singh, "A Survey of AI Text-to-Image and AI Text-to-Video Generators," IEEE, 2020. Saeedeh Parsaeefard, Iman Tabrizian, Alberto Leon-Garcia, "Artificial Intelligence as a Service (AI-aaS) on Software-Defined Infrastructure," IEEE, 2019.