Multi-domain Content Generation Using Advanced Machine Learning Techniques

Prajwal Urkude, Divya Charde, Shweta Pidurkar, Manjiri Dhavtode, and Mrudula Gudadhe

Abstract Innovation has a new meaning in our project as it involves the integration of cutting-edge technologies and generative models, thus making our stack of Next.js, React.js, Prisma ORM, MySQL and more scalable and easier to use for web development. Creative tools using AI like GPT-3.5 speedster and DALL-E 2 are used to bring ideas quickly that go beyond what has been done before with artistic work or music. Our multimedia capabilities have improved with Zeroscope-v2-xl, while an Expressive AI system captures authentic character nuances. Pinecone is used for vector databases, while Redis is used for efficient caching in order to make Llama-2-13b model mimic different personalities quickly and offer memory retention that is fast and long-lasting.

Keywords Generative models · GPT-3.5-turbo · DALL-E 2 · Artificial intelligence · Natural language processing · Langchain LLMs · Stable Diffusion · Zeroscope · Llama model

1 Introduction

There has been a lot of interest in generative technologies including AI-powered chatbots and automation tools for creative tasks in the past few years. A survey of more than 4000 AI professionals by Salesforce revealed that about half have used generative AI while one third use it regularly [1]. Examples like OpenAI's Midjourney, Dall-E and ChatGPT, as well as You.com's YouChat 2.0 [2] demonstrate how this technology can generate various creative outputs. As they process more data, generative AI models which utilize deep learning and neural networks tend to become better at producing genuine content. However, there is critical need for

P. Urkude (✉) · D. Charde · S. Pidurkar · M. Dhavtode · M. Gudadhe
Department of Information Technology, Priyadarshini College of Engineering, Nagpur, India
e-mail: it.prajwal.urkude@pcenagpur.edu.in

M. Gudadhe
e-mail: mrudula.gudadhe@pcenagpur.edu.in

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2024

J. Choudrie et al. (eds.), *ICT for Intelligent Systems*, Lecture Notes in Networks and Systems 1109, https://doi.org/10.1007/978-981-97-6675-8_45

proper supervision to ensure quality and accuracy because it can be invaluable in streamlining idea generation, content creation among other professional tasks [3].

Generative AI has introduced pivotal models for content creation. Generative Adversarial Networks (GANs) [4] comprise two neural networks—a generator that creates data and a discriminator that evaluates it. Variational Autoencoders (VAEs) [5] compress input data into a latent space and fabricate new data by sampling from this encoded representation. Transformer models enhance understanding of context in sequential data using neural networks, excelling in natural language processing tasks.

Our exploration is focused on changing generative AI by creating one consolidated system that combines modernized technologies. This merger encompasses web technologies, databases, and different AI models ranging from backend systems like Next.js and MySQL to ground-breaking AI tools like GPT-3.5-turbo and DALL-E 2. It makes a complete framework by introducing improvements from Pinecone and Redis as well as giving robust data management with MySQL and Supabase.

1.1 GPT-3.5-turbo

$$P(y_i|y_1, y_2, \ldots, y_{i-1}, x) = \frac{\exp(\text{score}(y_i, y_1, y_2, \ldots, y_{i-1}, x))}{\exp(\text{score}(yl, y_1, y_2, \ldots, y_{i-1}, x))}. \tag{1}$$

The formula calculates the likelihood of predicting token y_i based on previous tokens and input x. The softmax function scores each token option [6]. The diagram shows GPT-3.5-turbo's key components and workflow.

1.2 Dall-E

$$\mathcal{L}_{\text{VQ-VAE-2}} = Ez{\sim}q(z|x)\left[\left\|\,|sg[e](z)-z||_2^2\right\|\right]+\beta\|sg[d](x)-z\|_2^2. \tag{2}$$

The formula represents the loss function $\mathcal{L}_{\text{VQ-VAE-2}}$ for DALL-E 2, incorporating the reconstruction error [7] of the latent codes z and the commitment loss term with a hyperparameter β.

1.3 Stable Diffusion

$$\hat{p}_t = \frac{1}{Z_t} p_0(x_t) \Pi_{i=1}^{t-1} \frac{1}{Z_i} p_i(x_{i+1}|x_i). \tag{3}$$

The formula represents the estimated probability $\hat{p}_t$ of observing data x_t at time t in the Stable Diffusion model. It involves a product of conditional probabilities

$p_i(x_{i+1}|x_i)$ over previous time steps and initial probability $p_0(x_t)$. The normalization terms Z_t ensure that the probabilities form a valid distribution.

1.4 Zeroscope-v2-xl

$$\text{Attention(Q,K,V)} = \text{soft max}\left(\frac{\text{QK}^T}{\sqrt{d_k}}\right)\text{V}. \tag{4}$$

The formula represents the scaled dot-product attention mechanism used in Zeroscope-v2-xl. Given query Q, key K, and value V, the attention scores are computed by taking the dot product of Q and K divided by the square root of the dimensionality d_k. The softmax function is applied to obtain the attention weights, which are then used to weight the values V to produce the final attention output.

1.5 Llama-2-13b

$$\hat{y} = \arg\max_y P(y|x). \tag{5}$$

The formula represents the prediction of the most likely output ý given input x in the context of Llama-2-13b. The output is the token with the highest probability according to the model's prediction distribution $P(y|x)$, obtained through fine-tuning on large language models (LLMs) and vector storage mechanisms.

2 Literature Survey

Michael Cahyadi and colleagues evaluated DALL-E 2 and Luna [8], two image generation systems, using Google's Drawbench Benchmark. DALL-E 2 excelled over Luna in FID scores and human evaluations, though both struggled with concept generation and dataset biases. The authors suggest that expanding the model and dataset could enhance the accuracy and robustness of the system's performance.

Yunhe Feng's team explored the role of social media data in code generation with a GPT model [9], discussing its effectiveness and challenges. The paper details the process from data collection to GPT framework design and considers the ethical implications of using such data.

Gao et al. reviewed chatbot research [10], defining types and evaluating their limitations. They addressed ongoing issues in chatbot development, including data quality and ethics.

Sibo Zhang and associates presented a framework for generating talking-head videos from text [11], boasting a vast phonetic dictionary. Their method proved superior in visual fidelity and user preference across multiple languages.

Wu Ting studied computer-aided music generation, proposing a three-step composition algorithm that closely mimics original music [12]. The study suggests further enhancements in musical knowledge and feedback mechanisms.

Maryam Sadat Mirzaei and her team developed Animgan [4], a model for realistic human animations from images and movements, outperforming current methods on benchmarks like Human 3.6 M and recognizing areas for future improvement (Fig. 1).

3 Methodology

Datasets

GPT-3.5-turbo, trained on the CodeGenDataset [13], generates accurate and relevant code for a variety of programming tasks. DALL-E 2 creates high-quality images from the ImageGenDataset [14] by learning from an extensive collection of visual concepts. Stable Diffusion, utilizing the MusicGenDataset [15], produces expressive music pieces characterized by quality and diversity. Zeroscope-v2-xl, having learned from the VideoGenDataset [16], generates smooth and visually appealing video content. Llama-2-13b, employing the EmotionStyleReplicationDataset [17], accurately replicates character emotions and styles in its outputs.

Environmental Setup

React, a JavaScript library from Facebook, is designed for crafting user interfaces in single-page applications using components. It leverages a virtual DOM to streamline UI updates with diffing and reconciliation processes and supports unidirectional data flow for efficient state management. Tailwind CSS, created by Adam Wathan and others, is a utility-first framework that simplifies the development of custom interfaces with low-level utility classes and a just-in-time compilation strategy. Vercel maintains this open-source framework, which integrates with React for server-side rendering and static generation of web applications, aiming to provide scalable solutions for high-speed websites with hybrid rendering, easy routing, and serverless functions for backend integration. Shadcn UI enhances web development with customizable, user-friendly UI components built on Radix UI and Tailwind CSS.

Prisma, an open-source database toolkit for Node.js and TypeScript, offers an ORM system to facilitate backend connections with relational databases and provides developers with Prisma Client, Prisma Migrate, and Prisma Studio for data manipulation and management. Dataloader improves performance through batching and deduplication techniques, while MySQL organizes data using a relational model and SQL operations. PlanetScale delivers a scalable, secure MySQL-compatible service, handling tasks like patching and backups. Pinecone serves as a vector database for machine learning similarity searches, and Redis enhances data availability as

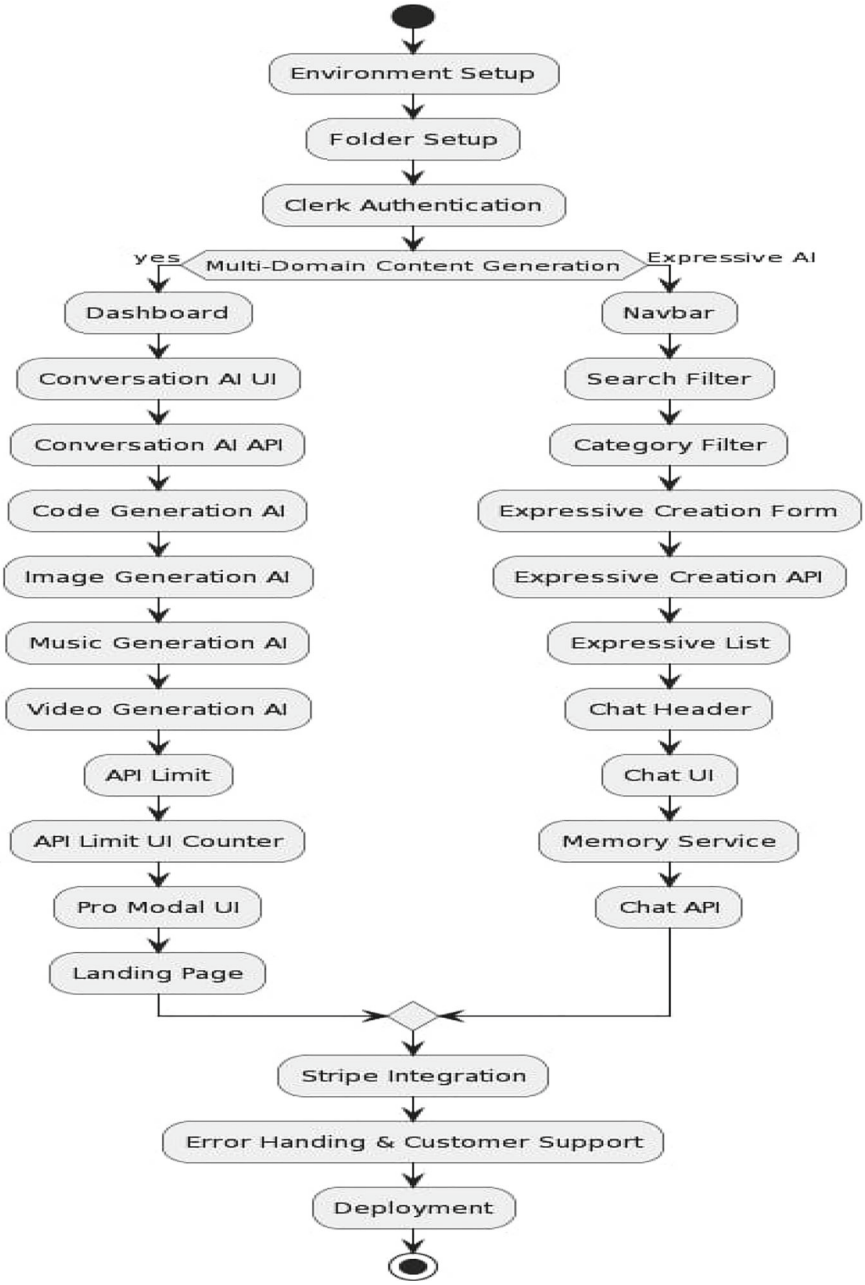

Fig. 1 Workflow/component diagram of project

an in-memory database, cache, and message broker. Cloudinary manages digital media assets with robust authentication and delivery mechanisms, and Crisp provides chat tools for immediate customer interactions. Stripe creates payment infrastructure for online transactions, and Clerk ensures secure web application authentication with OpenID Connect and OAuth 2.0, including multi-factor authentication. LangChain facilitates the development of applications using large language models with adaptable abstractions and AI-focused toolkits.

Database Configuration

The flowchart outlines a methodical process for setting up databases and caching in software applications, beginning with MySQL installation and schema setup, followed by Supabase project creation and PostgreSQL integration. It then guides through Pinecone account setup and SDK integration and concludes with Redis server configuration and cache policy establishment, encapsulating the entire configuration sequence.

Authentication Setup

Clerk provides web application authentication and user management, including registration, login, and profile services. It offers SDKs and APIs for easy integration with any tech stack and customizable front-end components for brand alignment. Install the Clerk SDK, integrate the provider, and use hooks and components for authentication. Clerk's useUser hook accesses user data, while SignIn and SignUp components allow for simple interface creation for user onboarding. A front-end API supports client–server communication.

UI Design

Shadcn UI incorporates Radix UI for functional components without predefined styles, promoting customization, and Tailwind CSS for utility-first CSS classes that expedite UI development. Accessibility is a priority, with components adhering to the Web Content Accessibility Guidelines (WCAGs) and including standard features like keyboard navigation and screen reader support.

Content Management: Set Up Cloudinary for Content Management

It is one of the initial steps to setup the content management system in our project. To utilize Cloudinary for content management, create an account and obtain your API key and credentials. Integrate the Cloudinary SDK into your project by following the installation instructions and initialize it with the API Key, API Secret, and Cloud Name.

API Keys: Develop APIs for Each AI Model to Facilitate Communication Between the Front-end and the AI Models

We need an OpenAI API key to integrate communication between the front-end and the AI models. Similarly, for Replicate API, secure the unique key post-registration, store it away from code, use it in the Authorization header for secure requests, and

ensure its safety by rotating and revoking as necessary to maintain secure service communication.

AI Models' Integration

Test models employ natural language processing and extensive language databases to execute a variety of tasks, including summarization, writing assistance, code generation, language translation, and sentiment analysis.

Conversation AI

Conversational AI benefits from an extensive range of training data, enabling it to comprehend and generate responses for various topics and user inputs. OpenAI's ChatGPT 3.5 Turbo [10] serves as a versatile language model for natural language processing tasks, excelling particularly in conversational AI applications. LLMs denote transformer language models that comprise hundreds of billions, or more, parameters and receive training on vast text datasets like GPT-3.

Code Generation

A code generator is designed specifically for producing code based on user prompts. It offers a focused solution that crafts task-specific code, beneficial for users seeking code snippets or solutions. Trained to understand and respond to user prompts, it adjusts its output to meet the coding requirements presented by the user. This software specializes in code generation, presenting a targeted and efficient solution for individuals requiring specific codes. Its primary function is as a referential model, yet it does not interact directly with other programs.

Image Generation

Dall-e 2 [8] generates realistic images that encapsulate the essence of the input. Such versatility allows for a broad spectrum of applications, ranging from artistic endeavors to functional image production. Users direct the model with prompts, influencing the style, content, and details of the generated images, culminating in a customized and context-sensitive image generation experience. This model stands as a potent instrument for both creative pursuits and practical imaging tasks.

Video Generation

In AI-driven video generation, Zeroscope v2 [18] marks a significant achievement. Zeroscope v2 empowers users to effortlessly create professional-grade videos from textual content. A hallmark of Zeroscope v2 is its user-friendly interface, accessible to both experts and novices. By altering video production methods and offering novel opportunities for businesses, content creators, and marketers, this interface, along with scalability and customization options, has transformed video creation.

Music Generation

Riffusion [19] is a refined library dedicated to real-time music and audio generation, employing Stable Diffusion techniques innovatively. This AI-powered platform translates textual input into lively and expressive musical compositions. Leveraging

artificial intelligence, Riffusion transforms written content into musical sequences, offering a distinctive blend of linguistic and auditory creativity. The stability in diffusion guarantees consistent and polished musical results, rendering it an essential resource for those seeking a dependable and sophisticated text-to-music AI tool.

Expressive AI

Llama-2-13b [20] is part of the Llama 2 series of advanced large language models developed by Meta. This series includes a variety of generative text models, both pre-trained and fine-tuned, with parameter counts ranging from 7 to 70 billion. The Llama-2-13b model contains 13 billion parameters and has been converted to the Hugging Face Transformers' format. The Llama 2 models excel in dialogue applications and have been fine-tuned to align with human standards for helpfulness and safety. This model utilizes supervised fine-tuning and reinforcement learning with human feedback to better conform to human preferences.

Testing

The chart presents a range of testing methodologies, from unit to automated testing tools, ensuring a thorough validation of generative AI tools for accuracy and reliability. Asterisk notations highlight the critical interplay between web technologies and testing methods. This flowchart epitomizes the intricate synergy of advanced AI systems within today's technological landscape (Table 1).

Customer Support

Install the Crisp SDK via npm, and configure it in a new crisp.js file within the components directory. Import the Crisp component using dynamic imports for runtime code loading and embed it in the application. Note that Next.js 13 projects may require converting class components to functional ones and adding 'use client' at the file's start.

Payment Gateway: Integrate Stripe for Payment Processing

All first, open an account with Stripe to get an API key and later incorporate Stripe SDK via npm. You can then set up a stripe.js file in the components directory for configuring it. With the expectation of different project requirements, we dynamically load the Stripe component.

Deployment: Deploy the Application on a Suitable Platform

The application undergoes development and testing before the developer transfers the code to GitHub and links it to Vercel. After setting up environment variables in Vercel, including API keys and database credentials, the developer initiates deployment. Vercel then connects with essential services such as PlanetScale for MySQL, Redis for caching, Supabase for additional database support, Cloudinary for content management, Crisp for customer support, and Stripe for payment processing. Each service validates its connection. Once all connections are established, Vercel confirms the successful deployment to the Developer, ensuring that the application is ready for user access with full functionality.

Table 1 Comprehensive table that includes learning rate, datasets, calculated accuracy, epochs, training data, models, max tokens input, and max tokens' output

Model	Learning rate	Dataset	Calculated accuracy (%)	Epochs	Training data	Max tokens input	Max tokens output
GPT-3.5-turbo	0.0001	CodeGenDataset	95	10,000	Code snippets and examples	512	1024
DALL-E 2	0.001	ImageGenDataset	92	5000	Diverse image dataset	256	512
Stable Diffusion	0.0005	MusicGenDataset	88	8000	Music compositions and genres	128	256
Zeroscope-v2-xl	0.0003	VideoGenDataset	96	12,000	Video clips and styles	512	1024
Llama-2-13b	0.0002	EmotionStyleReplicationDataset	98	15,000	Character emotions, styles, and memories	256	512

4 Conclusion

The paper claims that the project is complex due to incorporation of sophisticated technologies such as GPT-3.5-turbo for understanding words, DALL-E 2 for creating imaginary images, and Stable Diffusion for composing music. For video prediction, Zeroscope-v2-xl applies attention mechanisms, while llama-2-13b improves natural language processing through large model fine-tuning. In summary, these AI models improve text, image, music, video, and character rendering in showing a dedication to the progress of artificial intelligence and laying a foundation for future creativity.

5 Future Scope

The project is designed to bring about a stunning strategy of content generation for all platforms using artificial intelligence that produces music, videos, codes, and images including characters having advanced characteristics. In our forthcoming experiments, we shall include healthcare imaging, 3D object design, and aging faces to improve accuracy and robustness. Our goal is to make AI smarter at incremental learning and promote continuous improvement and adaptability for this purpose. This advantage can reduce costs, increase revenues, and free up time for mission-essential tasks. Additionally, we are building comprehensive datasets for different objectives. The developmental path of generative Al is so dynamic that ongoing research drives its way and greatly affects not only the society but also many sectors.

References

1. Salesforce, https://www.salesforce.com/news/press-releases/2023/09/07/ai-usage-research/. Last accessed 1 Dec 2023
2. Murugesan S, Cherukuri AK (2023) The rise of generative artificial intelligence and its impact on education: the promises and perils. Computer 56(5):116–121. https://doi.org/10.1109/MC.2023.3253292
3. TechRepublic, https://www.techrepublic.com/article/what-is-generative-ai/. Last accessed 1 Dec 2023
4. Mirzaei MS, Meshgi K, Frigo E, Nishida T (2020) Animgan: a spatiotemporally-conditioned generative adversarial network for character animation. In: IEEE international conference on image processing (ICIP), Abu Dhabi, United Arab Emirates, pp 2286–2290. https://doi.org/10.1109/ICIP40778.2020.9190871
5. Razavi A, Van den Oord A, Vinyals O (2019) Generating diverse high-fidelity images with vq-vae-2. In: Advances in neural information processing systems, vol. 32
6. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901
7. Kingma DP, Dhariwal P (2018) Glow: generative flow with invertible 1x1 convolutions. In: Advances in neural information processing systems, vol 31

8. Cahyadi M, Rafi M, Shan W, Lucky H, Moniaga JV (2023) Accuracy and fidelity comparison of Luna and DALL-E 2 diffusion-based image generation systems. In: IEEE international conference on Industry 4.0, artificial intelligence, and communications technology (IAICT), BALI, Indonesia, pp 108–112. https://doi.org/10.1109/IAICT59002.2023.10205695
9. Feng Y, Vanam S, Cherukupally M, Zheng W, Qiu M, Chen H (2023) Investigating code generation performance of ChatGPT with crowdsourcing social data. In: IEEE 47th annual computers, software, and applications conference (COMPSAC), Torino, Italy, 2023, pp 876–885. https://doi.org/10.1109/COMPSAC57700.2023.00117
10. Henno J, Jaakkola H, Mäkelä J (2023) What do we know about learning—conversations with chatGPT. In: 46th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 2023, pp 572–577, https://doi.org/10.23919/MIPRO57284.2023.10159818
11. Zhang S, Yuan J, Liao M, Zhang L (2022) Text2video: text-driven talking-head video synthesis with personalized phoneme-pose dictionary. In: ICASSP 2022–2022 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2659–2663. IEEE
12. Ting W (2020) Research on computer-aided music generation based on user-tag-media semantic mining. In: International conference on innovation design and digital technology (ICIDDT), Zhenjing, China, pp 17–22. https://doi.org/10.1109/ICIDDT52279.2020.00011
13. Hugging Face, gpt-3.5-turbo data from hugging Face, https://huggingface.co/datasets/wikitext. Last accessed 1 Nov 2023
14. Dall-e 2 data from paperswithcode, https://paperswithcode.com/dataset/lsun. Last accessed 2 Oct 2023
15. Stable diffusion data from Magenta, https://magenta.tensorflow.org/datasets/maestro. Last accessed 13 Oct 2023
16. Zeroscope-v2-xl data from paperswithcode, https://paperswithcode.com/dataset/kinetics. Last accessed 10 Sept 2023
17. Llama-2-13b data from Hugging Face, https://huggingface.co/datasets/AlekseyKorshuk/persona-chat. Last accessed 9 Nov 2023
18. Plugger.ai, https://www.plugger.ai/models/zeroscope-v2, Last accessed 2 Nov 2023
19. Stable diffusion for real-time music generation, https://replicate.com/riffusion/riffusion. Last accessed 10 Sept 2023
20. Llama-2–13b:- https://replicate.com/meta/llama-2-13b-chat. Last accessed 5 Nov 2023