Text To Image Synthisis Using Modified GANs

Dr. Shah S. N[1], Dr. Nazirkar S. B[2], Pokar Gaurav[3], Taware Swanand[4], Patil Nikhil[5], Suryawanshi Shruti[6]

[1]*Head of Department, Department of Computer Engineering*

[3456]*BE Students, Department of Computer Engineering*

Sharadchandra Pawar College of Engineering and Technology, Someshwarnagar, Pune, India

***Abstract*—Today, with the emergence of artificial intelligence and automated systems, the demand for tools that are capable of generating creative and intelligent content exponentially increases every day. Traditional methods of producing graphics and hand illustrated images take a great deal of time and effort; often requiring great technical expertise. To help solve this problem, we have created Promtify – Text-to-Image Generation System as a web based SaaS application developed using Java Spring Boot, REST APIs and AI-based picture generation models. Promtify allows users to create and generate quality realistic images from text descriptions entered into text prompts. By bridging the gap between creative thoughts and producing imaginative output on the web, Promtify provides an avenue for everyone's creativity to be realized digitally.**

***Keywords*—Ride-sharing, blockchain, smart contracts, WebSockets, decentralized applications, Ethereum, deterministic wallets.**

I. INTRODUCTION

The development of intelligent systems capable of understanding and visualising human thought has never been greater, especially with the advances in Artificial Intelligence (AI), automation, and the digital creation of images. Traditional image-creation methods are often lengthy and require a high level of ability or specialization to be successful.

To address this issue, Promtify has created an online service that allows for users to easily create images from text through the integration of AI, machine learning, and natural language processing (NLP). Promtify is accessed via a secure login system and uses credits to generate high-quality images directly from written descriptions.

Promtify is a Java-based Microsoft Platform as a Service (PaaS) using Spring Boot and is designed for secure user logins, image creation, generation of image credits, and recharging. Promtify supports the roles of Admin and User, thus facilitating the efficient management of image productions.

Promtify is easy to expand, modular and compatible with advanced AI models, which makes it the ideal product for designers, educators, marketers, and developers to quickly create digital content. Producing an image from a written description is a significant and difficult challenge for machine learning and involves resolving problems with both language modalities (among other things) by recognising how to control and manage partial or incomplete/ambiguous information using Natural Language Processing (NLP) techniques, as well as using other computer vision methods and techniques. This has now become one of the newest and most researched fields in computer vision. Generating an image from a text description is the reverse of creating a text caption or label from an image; therefore, text-to-image generation will provide the same kind of contextualization and relationship between the original description of the image and the final visual representation (image) as would you have got using the image captioning process. Furthermore, there are many applications in art, design, image retrieval, and searching for text-to-image generation. Currently, the majority of existing methods for generating images from a text description rely on a standard approach to creating a semantic representation of the text description using a pre-trained text encoder; once the semantic representations are completed, a Conditional GAN is used to train a Generative Adversarial Network to produce realistic face images.

II. PROBLEM STATEMENT

Current image creation approaches are slow, manual and require an extensive background in design. A few AI based image generators exist, but they tend to be very complicated, cost prohibitive, and not available to the general public. In addition, they often lack secure authentication, user control, credit based usage tracking, and the scalability needed to gain widespread usage.

Therefore, there is a need for a simple, intuitive, efficient and intelligent SaaS service based photo generation platform, using AI driven generation APIs, that produces images directly from users' given text descriptions, while providing users with real-time access to create, manage, and download those same images.

Promptify is the proposed photo generation system, and will solve the problems outlined above through a secure, credit based Software-as-a-Service (SaaS) system that has built-in access to AI image generation APIs, allowing users to create, manage, and download images easily and quickly in real time.

III. LITERATURE SURVEY

The conceptualization of Promptify is heavily influenced by a number of key academic articles in the area of Generative AI, Diffusion Models, or SaaS-based creative applications that have informed the system's development. Below is a synopsis of the two main literature sources contributing to the creation of Promptify:

1. P. Dhariwal et al. (2022) – "Diffusion Models Are BetterThan GANs" This article discusses diffusion models as an improved solution for image synthesis via Generative Adversarial Networks (GANs). Diffusion models produce highly realistic images by reducing noise from random samples through a probabilistic process. Diffusion models also handle issues with mode collapse and instability in training, which are characteristic of GANs. This model has been adopted in the design of numerous current AI-based image generation applications, including Stable Diffusion and DALL·E 2, both of which have made significant contributions to the development of Promptify through their influence.

2. A. Ramesh et al. (2022) – "CLIP Latent Space" UsingHierarchical Text-Tied to Image Generation (DALL·E 2) This paper provides a description of DALL·E 2, a model for generating semantically accurate images from text prompts. CLIP embeddings create a single latent space for the visual representation and the text prompt, ensuring semantic and creative alignment of the two modalities. This research directly influenced Promptify's design principles regarding the integration of language and images and prompt-based interaction with an AI."

3. R. Rombach et al., (2022) " High-resolution imagesynthesis through the use of latent diffusion models (LDMs)" – proposed a stable model of diffusion, which is a system that generates images based on a compressed latent space instead of a pixel-based space. This innovation leads to faster processing times as well as lower costs associated with computing since it uses less memory. This innovation will help greatly in achieving Promptify's goal of delivering real-time synthetic AI images using web APIs.

4. Y. Zhang et al., (2023) "Towards a controllable text-toimage generator by enhancing the understanding of prompts" This paper provides an analysis of how the AI understands the structure of a language in the prompt provided by a user and provides them with better control over the generated image in terms of Color/style/Composition.

Promptify provides users similar capabilities to customize the generated images so that they have greater artistic control and expressiveness.

5. K. Sun et al., (2023) "Commercialization of AI-poweredcreative tools as SaaS" This paper discusses how AI-powered creative tools can be commercialized using a software-as-aservice (SaaS) model. The key emphasis was placed on the importance of secure authentication, scaleable infrastructure, and credit-based systems for AI-based applications.

Promptify's credit management system and secure back-end design are based on the findings of this study. Promptify strives to provide a consistent, accessible and sustainable AI-based service.

IV. SYSTEM ARCHITECTURE

The three-tier architecture model of Promptify provides modularity, scalability, and security at each layer.

A. Frontend (Presentation Layer):

•Frontend built using HTML, CSS, JavaScript, Tailwind

•Provides users with a user-friendly interface to log in, add text prompts and images generated from them, and manage their credits and dashboards.

•Optimized for responsive and easy-to-use across a variety of devices.

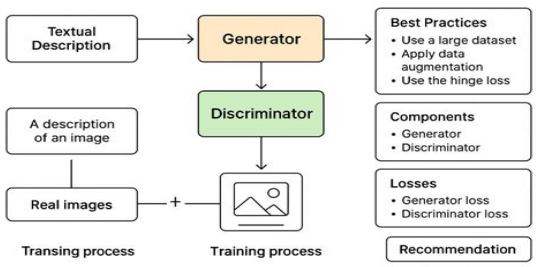

Fig 1: System Architecture

B. Backend (Application Layer):

•Backend was built using Java Spring Boot and handles the business logic for the permissions on each user, as well as managing the authentication and API integrations with AI systems.

- Handles credit-based access, user authentication, promptprocessing and communicating with AI image generation APIs (like Stable Diffusion or DALL·E).
- Implements Spring Security and JWT tokens for secureand encrypted data handling.
- Uses RESTful APIs for efficient retrieval of user data, prompt history, generated image URLs and transaction records.
- All sensitive information (passwords, tokens and paymentdata) is encrypted to ensure privacy and integrity.
- Regular backups, role-based access control and securestorage are supported.

V. SYSTEM MODULES

1) User Module- Handles user registration, logins, and credit use; Allows users to enter prompts, create and download images.
2) Admin Module- Manages user tracking, credits for plans, and overall system function; Contains analytical and activity logs for tracking; Provides for transparency and security.
3) AI Integration Module- Interface with external AI Models (Stable Diffusion, DALL·E, DeepAI, etc.) using RESTful services; Converts text prompt to images in real time.
4) Credit Management Module- Tracks credit use for the images created; Provides recharge of credits to users through payment processors (Razorpay, Stripe).
5) Security and Authentication Module- Manages user sessions, data encryption, and access control; All communication is done over HTTPS to protect against data breach.

VI. SYSTEM WORKFLOW

1) A user logs into the Promptify web application via a secure login
2) A user enters a prompt for the type of image they would like to create
3) The prompt is processed within the backend and makes a call to the AI model API for the generation of an image
4) The generated image will display on the frontend as well as be stored in the database
5) A user's credit balance will be decreased and a record of the transaction created
6) An administrator can observe all activities and manage users from the administrator dashboard

VII. PROPOSED SYSTEM:

The Proposed System - Promptify is a web-based platform that uses AI to produce truly outstanding images from text that have been created with cutting-edge technologies in the field of artificial intelligence (AI). The aim of this platform is scalable and secure, with a user interface that is designed to be easy to use and includes a credits system for earning / auto-generating / storing your own photography

A. Application Server (Back end):

1) Built with Java Spring Boot Framework for both reliability and modular design = Most of the server's main functions also include: User Sign in / Sign UP with Associated Account(s) Using (Spring Security, JWT), Prompt Management Credit Deductions/Recharge (As Required)

The main goal of the Proposed System: AI API Integration = This will allow secure integration of the Secure communication is guaranteed via RESTful APIs. • Create an automated conversion system from text to images that enables natural language prompts to be transformed into photorealistic images.

- Provide a secure, credits-based user environment withauthentication and authorization methods to control access.
- Integrate AI-based image creation APIs (e.g., DALL·E,Stable Diffusion) into the platform to allow for easy and accurate image creation.
- Allow administrators to monitor user activity, manageusers, and see usage analytics.

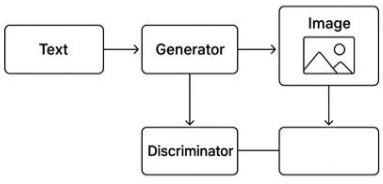

Fig 2: Block diagram of proposed system

B. System Description

The Promptify platform features a user-friendly frontend interface (or user interface), back-end application, and database connected to AI image generation APIs.

1) Front-End (User Interface):
   - Developed using HTML, CSS, JavaScript, and Tail-wind CSS - responsive user experience and ease of use. • Users are able to create logins, enter prompts, view images generated from their prompts, and see how many credits they have.
   - Simple and clear user interface for a better experience.• User Authentication: Using either encrypted credentials via JWT and secure login/registration.
   - Prompt Processing: Assuming text input and sendingthat input to the AI integrated APIs for image generation. • AI Image Generation: Interfaced with third-party API's (e.g. stable Diffusion, DALL·E 2) to create either realistic or artistic images.
2) Database (Data Storage Layer):
   - Either MySQL or Firebase Firestore is used to save:
   - User Profiles / authentication data
   - Historical data for prompts and URLs of generatedimages.
   - Credit usage and payment transactions.
   - Implemented encryption, and backing-up of all data.

C. Functional Features:

Mechanisms to protect user data.

- Credit Management System: Each user will have a stipu-lated dollar amount each time an image is generated by their entered prompt, deductions will be made from the users credit amount for that corresponding image generation. Users will be able to recharge their accounts via Razorpay / Stripe payment gateways.
- Admin Dashboard: Administrators will be able to monitoruser activity, manage user credits, track usage data, and generate performance reports for the platform.
- Data Security: All platform communications will be overHTTPS, and all sensitive data will be securely encrypted, and role-based access control to platform systems.

D. Advantages of Proposed Solution

- Automation: No manual efforts required to create digitalimages.
- Easy to use online web interface accessible from anydevice.
- Multiple layers of security (encryption, JWT, HTTPS).
- Supports cloud deployment and future growth due to itsmodular architecture.
- Minimal resources are used to produce images quicklywith answers you want to see.
- Affordable and easy access due to a credit-based pricingmodel.

E. Expected Results

- A working AI powered image generation platform that cancreate high-quality images from text-based prompts.
- Reliable and secure management of users, prompts andcredits.
- A better way for creatives of all types (designers, devel-opers, educators, marketers) to be able to create using visual imagery.
- A basis for expanding future options such as customizingstyles, multilingual prompts and creating 3D content.

VIII. RESULT

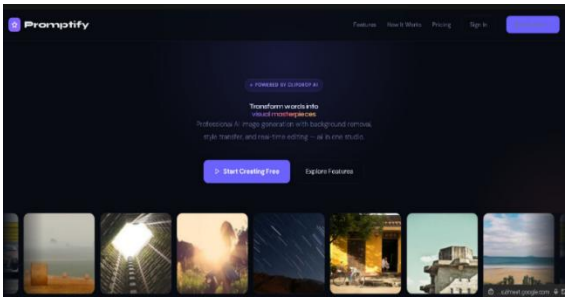

Fig 3: Login page

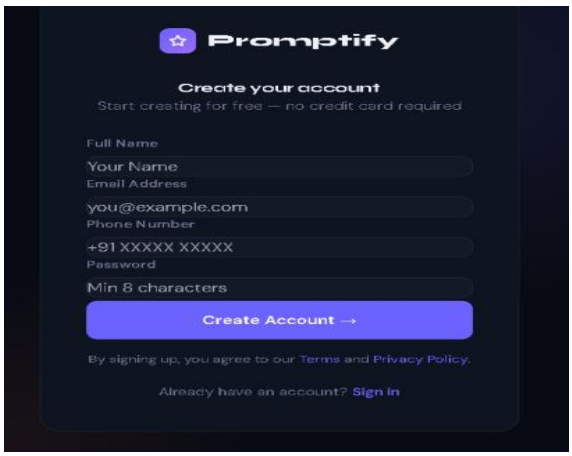

Fig 4: Account Create

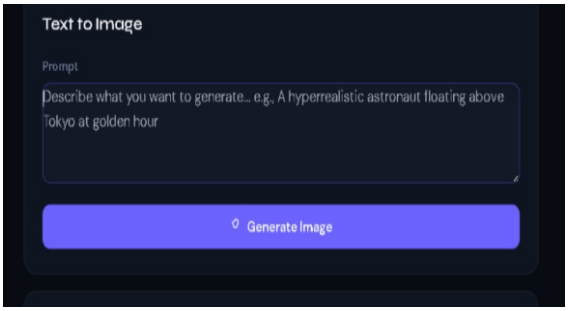

Fig 5: Description for generate image

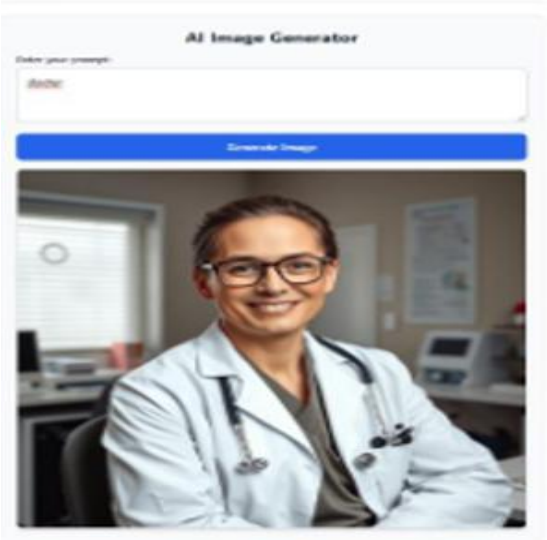

Fig 6: Generate image

IX.ALGORITHM

The proposed algorithm can be summarized as follows:

1. Input: User enters a text prompt describing the desired image.
2. Prompt Processing: The system parses and analyzes the prompt using natural language processing (NLP) to extract context, style, and key objects.
3. Model Selection: The backend selects an appropriate AI model (e.g., Stable Diffusion or DALL·E) based on the prompt complexity and parame ters.
4. Image Generation: The model converts the processed prompt into a high-quality image using deep learning and diffusion techniques.
5. Post-Processing: The generated image is refined, upscaled, and optimized for clarity and resolution.
6. Output: The final image is displayed to the user through the web interface and optionally stored in the database for download or sharing.

X. CONCLUSION

The Promptify – Text-to-Image Generation System exemplifies combining artificial intelligence with web developments to convert simple text descriptions into high-quality, realistic photos.

The Promptify system is capable of providing a userfriendly, scalable, and robust approach to providing a platform for digital creativity through its integration of Java Spring Boot technology for the back end; RESTful APIs to communicate between the different system components; and popular AIbased image generation models such as Stable Diffusion and DALL·E.

There are two primary objectives of this project:

- To give users a means for automating the creation ofimages through the application of artificial intelligence.
- To provide users with a secure, efficient method forauthenticating themselves and accessing the service on a credit-based system.

The modular nature of the Promptify system supports the maintainability, scalability, and security of user data while the use of a credit-based system increases the economic viability and accessibility of the solution.

In summary, Promptify demonstrates the gap that exists between human imagination and machine intelligence in the world of AI-powered creative technology.

Future Scope

- The addition of style and color customizable tools.
- The ability to process prompts in multiple languages toprovide for wider accessibility.
- Optimization of image generation on low-end devices,yielding faster image generation times.
- Expanded real-time editing capabilities and features topermit the generation of 3D or video illustrations.

With continued development, the Promptify solution could potentially provide a leading platform for AI-enabled content creation across the design, marketing, education, and entertainment industries.

REFERENCES

[1] Dhariwal, P., et al. (2021), Diffusion Models Outperform GANs, Advances in Neural Information Processing Systems, NeurIPS.

[2] Ramesh, A., et al. (2022), CLIP Latents for Hierarchical TextConditioned Image Generation (DALL·E 2) at OpenAI.
[3] Rombach, R. et al. (2022), High-Resolution Image Creation Using Latent Diffusion Models, in the Proceedings of the 2022 Computer Vision and Pattern Recognition Conference.
[4] Zhang, Y., et al. (2023), Towards Controllable Text-to-Image Generation Through Enhanced Prompt Understanding, IEEE Transactions on Neural Networks and Learning Systems.
[5] Sun, K. and others. (2023), AI Based Creative Tools Delivered via SaaS, AIigament for SaaS Solutions for Cloud Computing.
[6] H.Zhang, T.Xu, H.Li, S.Zhang, X.Wang, X.Huang, and D.N.Metaxas, StackGAN: Realistic image synthesis with stacked generative adversarial networks, IEEE Transactions Pattern Analysis and Machine Intelligence, Volume: 41, Issue: 8, Aug. 2019, Pages: 1947-1962.
[7] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, The Caltech UCSD Birds-200-2011 Dataset, California Institute of Technology, Pasadena, California, USA, Tech Report, 2011.
[8] M.-E. Nilsback and A. Zisserman, Automated flower classification across a large number of categories, in Proc. 6th Indian Conference on Computer Vision, Graphics and Image Processing, Dec. 2008, Pages: 722-729.
[9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, C.L. Zitnick, Microsoft COCO: Common Objects in Context, in Proc. European Conference on Computer Vision, Cham, Switzerland: Springer, 2014, Pages: 740-755.
[10] D.P. Kingma and M.Welling, Auto-Encoding Variational Bayes, 2013, arXiv:1312.6114, [Online]. Retrieved from http://arxiv.org/abs/1312.6114.