IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI Driven Artwork Generation using Text Description

Janhavi Sachin Sonaje[1], Ashwini Jagdish Patil[2], Kanchan Devidas Patil[3], Nikita Anil Sonawane[4], Dr. Rajnikant Wagh[4]

[1]Dr. Rajnikant Wagh, Professor :0000- 0003- 2997-6034
R. C. Patel Institute of Technology , Shirpur, India, 425405

Abstract *– This research presents an interactive platform for AI-driven artwork generation, transforming written text into expressive, high-quality visuals for users of all artistic skill levels. The system integrates a pretrained Stable Diffusion model with a robust MERN stack (MongoDB,Express.js, React.js, Node.js) web application. This architecture provides an accessible digital canvas that instantly turns ideas into visuals, making it a powerful tool for rapid ideation, storytelling, and accelerating creative workflows for professional designers and artists. Key features include support for high-resolution output and a categorized style selection, such as Meme, 3D Cartoon, 3D illustration, Sketch, Ghibli, Abstract and Logo generation enabling precise artistic control. Ultimately, the system enhances productivity while preserving the uniqueness of each user's vision, offering a vital bridge between imagination and innovation.*

Keywords- *Style Selection, Stable Diffusion, Generative AI, MERN stack, Artistic skills, high quality visuals*

1. INTRODUCTION

In recent times, artificial intelligence (AI) has moved beyond its traditional operations in data processing and analytics, entering disciplines of creativity such as art, music, and design [1]. One of the most transformative inventions in this space is text-to-image generation, which enables users to produce images directly from natural language descriptions. This technology represents a significant step in standardizing creativity, empowering individuals without artistic training to transform their imagination into visual output.

Despite the availability of tools such as DALL·E, Midjourney, and Bing Image Creator existing solutions often come with challenges, including limited usage quotas, subscription requirements, and limited customization options [1]. These constraints can hamper accessibility, particularly for learners, researchers, and independent creators. Additionally, many commercial platforms serve as "black boxes" offering little transparency into the underlying AI models, reducing their value as educational resources [2].

To address these gaps, this exploration presents an interactive, open-access platform that transforms textual input into suggestive, high-quality digital artwork. The system is built by integrating a pre-trained Stable Diffusion model with a MERN stack(MongoDB, Express.js, React.js, Node.js) web application, enabling real-time image generation through a user-friendly interface. By doing so, it offers a seamless "digital canvas" where ideas are instantly visualized, lowering the barrier to entry for creative exploration.

A crucial point of this platform is its categorization of generated images into distinct artistic styles, including Meme Creator, 3D Image, Cartoon, Sketch, Logo, Ghibli, and Synaesthesia Generator. These categories allow users to align outputs with their intended purpose whether for concept art, professional design workflows, or casual creative experimentation [3]. Additionally, the platform supports high-resolution outputs and customizable prompts, granting users finer control over the artistic direction and resulting imagery.

By bridging human imagination with AI-powered art, this system contributes to both the creative technology ecosystem and the academic study of generative models. It provides a practical, accessible, and scalable solution for visual creativity, storytelling, and design prototyping, unlocking new opportunities for individuals and industries alike [4].

2. LITERATURE REVIEW

Early explorations in AI-generated art focused on the evolution from algorithmic creativity to generative adversarial approaches. Tools like Google Colaboratory facilitated experimentation with machine learning models, while surveys highlighted the transition from classical text-to-image models to diffusion-based architectures. Hugging Face transformers provided state-of-the-art NLP capabilities, and the introduction of Stable Diffusion enabled efficient, high-quality text-to-image synthesis [5]. Research on foundation models emphasized both opportunities and risks of large-scale AI systems, supported by Hugging Face Diffusers and Google Colaboratory integration for scalable deployment .
Studies examining human perception of AI-generated art revealed challenges in distinguishing AI from human-created works, prompting ethical frameworks for copyright, attribution, and user control.[5] The intersection of art and generative AI demonstrated creative applications, as well as economic implications in workplace environments. Legal considerations explored fair learning and copyright issues, while large-scale text-to-image models influenced visual artists' creative practices [6]. Analyses of online discussions further explored public perception of AI-generated art , and applications like Midjourney highlighted AI's role in artistic and advertising creativity.

Recent developments include AI-assisted art appreciation via text generation, step-by-step text-to-painting generators like SSG-Net, and comparative studies of AI-generated versus human art.[7] CNN-RNN encoder-decoder models enabled automated artwork descriptions, and AI image generators influenced product design. Research also addressed user authorship in AI-generated works , and textual analyses extended to AI-generated biblical art

3. METHOLOGY

The methodology of the system integrates both software engineering practices and AI-driven generative modeling to provide an interactive and user-friendly artwork creation platform. A JSON-based request structure is used for processing prompt submissions.

and generation tasks [16]. Each request includes session identifiers, timestamps, and user metadata to ensure traceability and organized logging throughout the system.

```
{

"prompt ": "Ghibli−style  landscape with  floating  islands " ,

"negative_prompt ": " distorted  shapes ,  low  quality " ,

" style ": "Ghibli " ,

"timestamp ": "2025−10−12T09:00:00Z" ,

"user_id ": "USER001"

}
```

3.1 Text-to-Image Generation

The Stable Diffusion XL (SDXL) model is employed for converting textual prompts into high-quality images. Positive and negative prompts refine model outputs and ensure alignment with user intent [16]. The Euler Discrete Scheduler is applied to optimize the sampling process, improving efficiency, stability, and clarity in the generated images. The backend automatically queues requests and retries them when server resources or GPU endpoints are temporarily unavailable [17].

3.2 Style Adaptation

To ensure flexibility in artistic output, the system supports multiple predefined styles, including Meme, Cartoon, 3D Illustration, Abstract, Ghibli-style, Logo and Live Image generation. Each style preset modifies the prompt structure, guiding the diffusion model to maintain consistency in visual characteristics across different requests[18].

3.3 System Deployment

The system is deployed using Google Colab for GPU-accelerated inference. FastAPI functions as the backend service layer, exposing inference routes for prompt processing and model execution. Ngrok is used to establish secure tunneling between the remote GPU environment and the application frontend, eliminating the need for a dedicated physical server while ensuring stable connectivity.

3.4 User Interaction and Storage

The frontend is designed using React.js to deliver dynamic interaction and realtime visualization of generated outputs. MongoDB acts as the primary database for storing generated images, prompt history, user sessions, and metadata [18]. Cloudbased services such as Cloudinary are used for image hosting and content delivery.
A structured logging mechanism links each prompt, timestamp, and resulting image. Redis caching optimizes repeated prompt retrieval and reduces redundant model execution, ensuring faster system responsiveness [19]. The storage architecture supports prompt reuse, historical browsing, and scalability for larger datasets.

3.5 Scalability and Flexibility

The system is architected with modularity to support migration from prototype to production environments. It can transition seamlessly from Google Colab to cloud platforms such as AWS, GCP, or Azure without major architectural changes [20]. Key scalability features include:

- Cloud Ready: Supports containerized deployment across cloud GPU platforms.
- Horizontal Scaling: Enables load balancing across multiple GPU instances for high traffic [21].
- Resource Optimization: Auto-scaling mechanisms ensure cost-effective GPU usage.
- Modular Architecture: Frontend, backend, caching layer, and model components can be updated independently.
- Caching and Load Management: Redis caching reduces model reload overhead and improves API response time.
- Asynchronous Processing: The API layer handles multiple simultaneous requests without performance degradation.
- Fault Tolerance: Retry mechanisms and failover support maintain uninterrupted service during network or node failures.
- Monitoring and Logging: Tools such as Prometheus and Grafana track GPU utilization, API latency, and user activity for proactive scaling.

3.6 Software Performance Matrics

Table 1 shows all the performance metrics of the proposed system

Table 1. Software Performance Metrics

Metrics	Value
Average Server Response Time	Approximately 200 milliseconds
Database Uptime (MongoDB)	99.9%
Dashboard Availability	99.9%

4. DESIGN

The proposed system is designed using a modular architecture that integrates modern web technologies with advanced deep learning models to generate artwork from textual prompts. The architecture ensures scalability, responsiveness, and efficient collaboration between client, server, and AI services [8].
At the core of the architecture lies the MERN stack (MongoDB, Express.js, React.js, Node.js), which facilitates seamless interaction between users and the backend services. A pre-trained Stable Diffusion model, hosted on Google Colab with GPU acceleration, serves as the AI engine for image generation [8]. FastAPI is employed to expose the model as a RESTful API, with Ngrok providing secure tunneling for remote access.

4.1 System Components

- Frontend (React.js): Provides the user interface where prompts are entered, styles are selected (e.g., Meme, 3D Cartoon, Ghibli, Sketch, Logo), and generated images are displayed.
- Backend (Node.js/Express.js): Acts as middleware, handling prompt validation, user requests, and communication with the FastAPI service.
- AI Service (FastAPI + Stable Diffusion): Executes text-to-image generation using the Stable Diffusion XL model. Runs in Google Colab with GPU resources for high-performance inference [9].
- Database (MongoDB): Stores user data, prompts, and metadata of generated artworks, enabling persistence and personalization.
- Storage/Delivery (Cloudinary): Optimizes image delivery via CDN for faster user access.

4.2 Stable Diffusion XL

Stable Diffusion XL (SDXL) represents a significant advancement over the base Stable Diffusion model, offering improved image quality, better composition, and more detailed outputs [10]. SDXL features a larger U-Net backbone with more parameters and employs a two-stage base and refiner model architecture for enhanced generation capabilities.

4.3 Key Components in SDXL

- Enhanced Architecture: SDXL uses a larger U-Net with approximately 2.6 billion parameters compared to the original 860 million, enabling better understanding of complex prompts and generating more coherent images [10].
- Two-Stage Generation: The model employs a base model for initial image generation followed by a refiner model that adds fine details and improves overall image quality, especially in textures and fine structures.
- Better Text Understanding: Improved text encoder and cross-attention mechanisms allow SDXL to better comprehend complex prompts with multiple subjects, attributes, and relationships [11].
- Higher Resolution Output: SDXL natively supports 1024×1024 resolution generation with improved sharpness and fewer artifacts compared to previous versions.

4.4 SDXL in Proposed System

The proposed system utilizes SDXL as its primary generation engine, leveraging its enhanced capabilities to produce high-quality images across various styles including Meme, 3D Cartoon, Ghibli, Sketch, and Logo generations [12]. The larger model size and improved architecture make it particularly suitable for generating diverse artistic styles while maintaining prompt adherence and visual quality.

4.5 Performance Considerations

Despite its larger size, SDXL maintains reasonable inference times through optimizations such as:

- GPU Acceleration: The system leverages NVIDIA T4 or V100 GPUs available through Google Colab, providing the parallel processing power necessary for efficient tensor operations and matrix computations fundamental to the diffusion process [13].
- Efficient Memory Management: Techniques like gradient checkpointing and optimized attention mechanisms reduce VRAM consumption by selectively recomputing activations during the backward pass rather than storing all intermediate values, enabling larger batch sizes and faster processing.
- Optimized Latent Space Operations: By performing denoising in a compressed latent space approximately 64x smaller than the original pixel space, SDXL significantly reduces computational complexity while preserving semantic information critical for high-quality image generation [14].
- Batch Processing Capabilities: The model architecture supports parallel processing of multiple prompts, improving throughput and enabling efficient resource utilization when handling concurrent user requests in production environments.

4.6 Workflow

Figure 1 shows system work flow which begins with the user entering a text prompt and selecting the desired style of artwork. The frontend sends this input to the backend, which forwards it to the FastAPI server via Ngrok [15]. The server processes the request and generates an image using the Stable Diffusion model. Finally, users can choose to save the generated image and metadata in MongoDB for future reference.

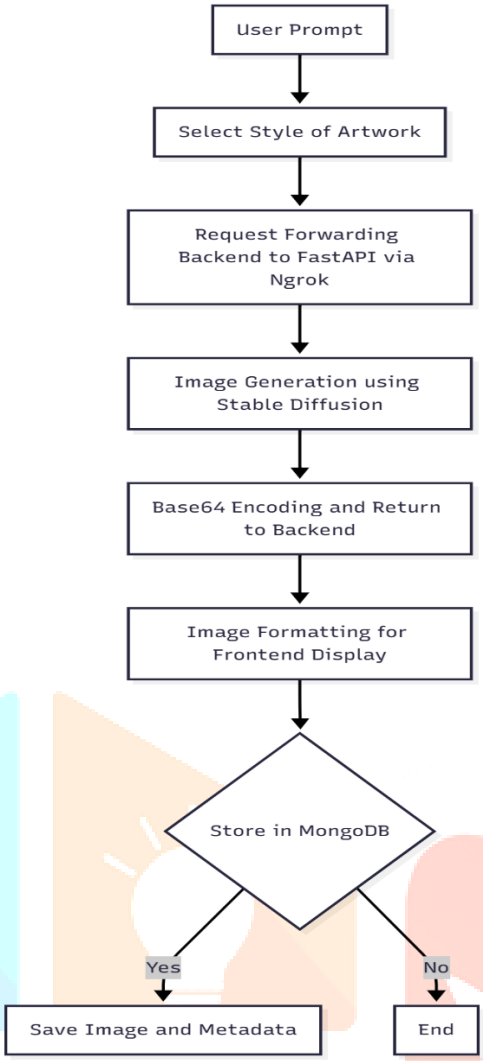

Fig. 1. Flowchart of system workflow

RESULT

4.7 Generated Image Output

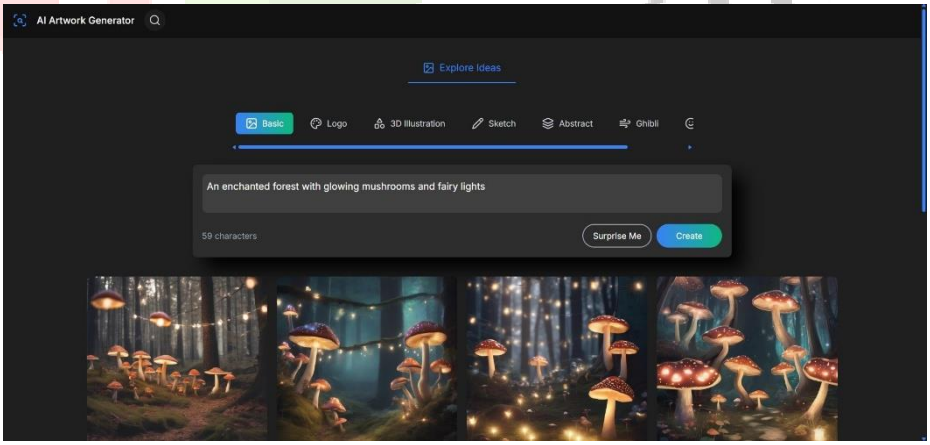

Fig. 2**.** Basic Image Generation

Figure 2 demonstrates the system's ability to generate high-quality images from textual prompts. The user selects a style from the style bar and enters descriptive text. The Stable Diffusion model produces four images in real-time according to the chosen style.

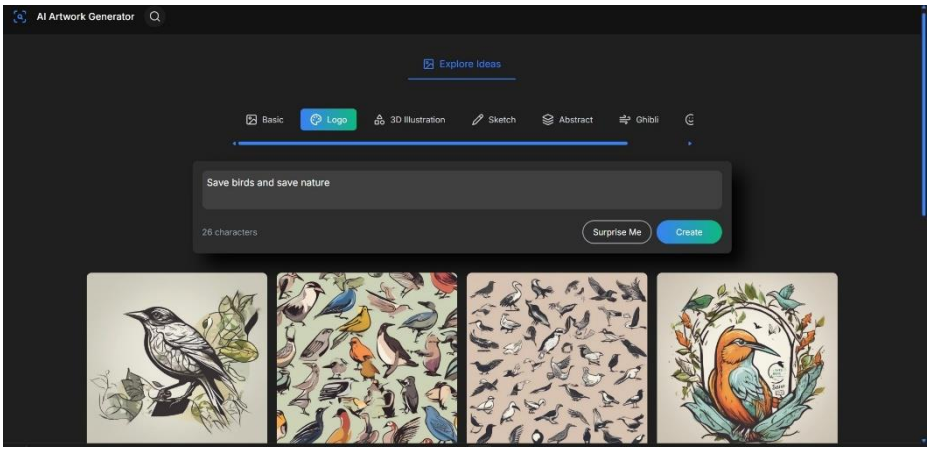

Fig. 3. Logo Generation and Display

Figure 3 illustrates the system's capability to generate logos based on user prompts. Generated logos are displayed immediately on the dashboard and can be downloaded or reused for subsequent design tasks.

CONCLUSION

The developed platform successfully bridges the gap between creative imagination and visual output by providing an accessible and user-friendly text-to-image generation tool. By leveraging a Stable Diffusion model and a MERN stack, the system not only serves as a practical tool for various applications from education to therapy but also as an open-source educational resource for students and developers.The project demonstrates how modern AI can be used to empower individuals, regardless of their artistic skill, and make creativity more inclusive. Future work could involve upgrading to paid cloud instances for higher throughput and exploring more advanced image generation pipelines. The platform highlights the scalability and flexibility of integrating deep learning models with web technologies, enabling real-time interaction and seamless user experience. The modular architecture ensures that new features and models can be easily incorporated without disrupting the existing system. It also promotes experimentation, allowing users to explore different prompts, styles, and configurations to achieve desired outputs.

REFERENCES

1. *Bisong, E.: Google Colaboratory. In:* Building Machine Learning and Deep Learning Models on Google Cloud Platform, *pp. 59–64. Apress, Berkeley, CA (2019).*
2. *Zhang, A., Qian, L.: AI Decide Artist: Text-to-Image Generation and Transformers. Preprint (2020).*
3. *Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Rush, A.M.: Huggingface's transformers: State-of-the-art natural language processing.* arXiv preprint arXiv:1910.03771 *(2020).*
4. *Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S.: On the opportunities and risks of foundation models.* arXiv preprint arXiv:2108.07258 *(2021).*
5. *Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In:* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), *pp. 10684–10695 (2022).*
6. *Bosonogov, S., Suvorova, A.: Perception of AI-generated art: text analysis of online discussions.* Zapiski Nauchnykh Seminarov POMI ***529****, 6–23 (2023).*
7. *Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M.R., Groh, M., Akten, M.: Art and the science of generative AI.* Science ***380****(6650), 1110–1111 (2023).*
8. *Brynjolfsson, E., Li, D., Raymond, L.R.: Generative AI at work. National Bureau of Economic Research Working Paper No. 31161 (2023).*
9. *Lemley, M.A., Casey, B.: Fair learning.* Texas Law Review ***101****(4), 743–782 (2023).*
10. *Samuelson, P.: Generative AI meets copyright.* Science ***381****(6654), 158–161 (2023).*
11. *Doctorow, C.: The generative AI revolution: opportunities and challenges for the creative industries.* Journal of Cultural Economics ***47****(2), 145–167 (2023).*
12. *Hanna, D.M.: The use of artificial intelligence art generator "Midjourney" in artistic and advertising creativity.* Journal of Design Sciences and Applied Arts ***4****(2) (2023).*
13. *Hedges, K.E.: Artificial intelligence (AI) art generators in the architectural design curricula. In:* 2023 ASEE Annual Conference Exposition, *pp. 1–14 (2023).*
14. *Ko, H.-K., Park, G., Jeon, H., Jo, J., Kim, J., Seo, J.: Large-scale text-to-image generation models for visual artists' creative works. In:* Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23), *pp. 919–933 (2023).*
15. *Das, M., Fiannaca, A.J., Morris, M.R., Kane, S.K., Bennett, C.L.: From provenance to aberrations: image creator and screen reader user perspectives on alt text for AI-generated images. In:* CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, *pp. 1–21 (2024).*
16. *Dartey, C.A.A., Sokolova, A., Martellini, A.: Generating descriptions of artworks with a CNN–RNN encoder-decoder model. Preprint (2024).*
17. *Gobbi, A.: Code, canvas, and copyright: establishing user authorship in AI-generated art. Preprint (2024).*
18. *Hermanto, Y.A.L.: Artificial intelligence-image generator (A.I.-I.G.): characteristics and influence in industrial development of design product. Taylor & Francis Online (2024).*
19. *Khan, F.F., Kim, D., Jha, D., Mohamed, Y., Chang, H.H., Elgammal, A., Elliott, L., Elhoseiny, M.: AI art neural constellation: revealing the collective and contrastive state of AI-generated and human art. In:* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, *pp. 7470–7478 (2024).*
20. *Rueda-Arango, Y.D., Rojas-Velazquez, D., Gorelova, A.V., Lopez-Rincon, A.: Exploring human perception of AI-generated artworks. In:* 2024 IEEE International Symposium on Technology and Society (ISTAS), *pp. 1–7 (2024).*
21. *Shende, T.: AI-enhanced art appreciation: generating text from artwork to promote inclusivity. In:* Proceedings of the 38th AAAI Conference on Artificial Intelligence *(2024).*
22. *Zhao, Y., Li, Y.: SSG-Net: a novel text-to-painting AI artwork generator enables illustrating of the step-by-step painting process. In:* Proceedings of the IEEE Ninth International Conference on Multimedia Big Data (BigMM) *(2023).*
23. *Makimei, H., Wang, S., van Peursen, W.: Seeing the words: evaluating AI-generated biblical art.* arXiv preprint arXiv:2504.16974 *(2025).*