

**Image Generation using stable diffusion & Comfy UI**

A Project Report

submitted in partial fulfillment of the requirements of

AICTE Internship on AI: Transformative Learning with

TechSaksham – A joint CSR initiative of Microsoft & SAP

by

**Dhruba,** [**dhrubanpanda.72@gmail.com**](mailto:dhrubanpanda.72@gmail.com) **STU6570696e6a64c1701865838**

Under the Guidance of

**Aditya Prashant Ardak**



### ACKNOWLEDGEMENT

I am immensely grateful for the guidance, encouragement, and unwavering support that I have received throughout the course of this project. This accomplishment would not have been possible without the contributions of several individuals and organizations, to whom I owe my deepest gratitude.

First and foremost, I wish to express my heartfelt thanks to my esteemed guide, **Aditya Prashant Ardak & Jay Rathod** for their exceptional mentorship and dedicated guidance. Their insightful advice and constructive feedback have been invaluable throughout every stage of this project. The patience and expertise demonstrated by my guide in addressing challenges, providing innovative solutions, and steering me in the right direction have been the cornerstone of this successful endeavor. Their belief in my abilities and continuous motivation inspired me to push the boundaries of my potential and achieve excellence in this work.

I am also profoundly grateful to **TechSaksham** for providing such an enriching platform to explore and implement innovative ideas in the field of artificial intelligence. The transformative learning experience and access to invaluable resources provided through this initiative have significantly enhanced my technical knowledge and professional development. The internship offered me a unique opportunity to translate theoretical concepts into practical applications, and I deeply appreciate the vision of TechSaksham in empowering young minds like me.



### ABSTRACT

This report explores the use of **Stable Diffusion and ComfyUI** for generating high-quality AI-driven images, focusing on procedural content generation and artistic visualization. Traditional image creation methods often require extensive manual effort, limiting scalability and customization. By leveraging **Stable Diffusion’s deep learning models** and **ComfyUI’s node-based workflow**, this project enables users to generate detailed and highly customizable images with minimal manual intervention.

The workflow integrates **ControlNet for structured composition, inpainting for refinement, and LoRA models for style adaptation**, ensuring precise control over artistic and photorealistic outputs. Through parameter tuning and node-based editing, users can create complex scenes while maintaining flexibility in image design.

This study highlights the efficiency of AI-driven image generation in fields such as digital art, concept visualization, and automated content creation. Future advancements will focus on **enhancing model fine-tuning, improving UI intuitiveness, and integrating real-time generation techniques**. The project demonstrates how generative AI can revolutionize the creative process, providing scalable and efficient solutions for diverse artistic and practical applications.



**TABLE OF CONTENT**

[Abstract I](#_bookmark0)

Chapter 1. Introduction 1

* 1. [Problem Statement 1](#_bookmark1)
  2. [Motivation 1](#_bookmark2)
  3. Objectives 2
  4. [Scope of the Project 2](#_bookmark3)

Chapter 2. Literature Survey 3

* 1. [Review relevant literature 3](#_bookmark4)
  2. [Existing Models, Techniques, and Methodologies 4](#_bookmark5)
  3. [Limitations in Existing Systems 5](#_bookmark6)

Chapter 3. Proposed Methodology 6

* 1. [System Design 6](#_bookmark7)
  2. [Requirement Specification 8](#_bookmark8)

Chapter 4. Implementation and Results 9

* 1. [Snap Shots of Result 9](#_bookmark9)
  2. [GitHub Link for project 14](#_bookmark10)

Chapter 5. Discussion and Conclusion 15

* 1. [Future Work 15](#_bookmark11)
  2. [Conclusion 17](#_bookmark12)

[References 18](#_bookmark13)



**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
| **Figure 1** | System workflow diagram for AI-based image generation using Stable Diffusion & ComfyUI. | **7** |
| **Figure 2** | Snapshot of first AI-based image generation using Stable Diffusion & ComfyUI | **9** |
| **Figure 3** | First AI-based generated image using Stable Diffusion & ComfyUI | **10** |
| **Figure 4** | Snapshot of second AI-based image generation using Stable Diffusion & ComfyUI | **11** |
| **Figure 5** | Second AI-based generated image using Stable Diffusion & ComfyUI | **12** |
| **Figure 6** | Snapshot of third AI-based image generation using Stable Diffusion & ComfyUI | **12** |
| **Figure 7** | Third AI-based generated image using Stable Diffusion & ComfyUI | **13** |
| **Figure 8** | Snapshot of forth AI-based image generation using Stable Diffusion & ComfyUI | **13** |
| **Figure 9** | Snapshot of fifth AI-based image generation using Stable Diffusion & ComfyUI | **14** |

# CHAPTER 1



## Introduction

#### Problem Statement:

methods Traditional methods of image generation and editing rely heavily on manual design, extensive artistic expertise, and time-consuming processes. Creating high-quality visuals requires professional tools, intricate workflows, and significant effort, making it difficult for non-experts to generate detailed and customized images efficiently. Additionally, manual editing can be error-prone and lack flexibility, especially when fine-tuning complex compositions.

An AI-driven image generation system using **Stable Diffusion and ComfyUI** offers an **automated, flexible, and efficient** solution. By leveraging **deep learning-based text-to-image synthesis**, users can generate high-quality images with fine-tuned control over style, composition, and details—eliminating many of the challenges associated with traditional methods.

**Why is this significant?**

**Efficiency:** Reduces the time and effort required for creating high-quality images.

**Customization:** Allows for precise control over artistic elements using node-based workflows.

**Accessibility:** Enables both professionals and beginners to generate complex visuals with minimal effort.

**Scalability:** Supports a wide range of use cases, from concept art to marketing and content creation.

#### Motivation:

This project was chosen due to the growing demand for AI-powered content generation across various industries. As artificial intelligence continues to revolutionize creative fields, leveraging tools like Stable Diffusion and ComfyUI provides an innovative solution to streamline image creation, design prototyping, and artistic exploration.

**Potential Applications and Impact:**

**Digital Art & Concept Design** – Automates concept generation for artists, game developers, and illustrators.

**Marketing & Advertising** – Generates customized visuals for branding, promotions, and social media campaigns.

**Film & Animation** – Assists in storyboarding and environment design for movies and animation.

**E-Commerce & Product Design** – Creates photorealistic product mockups and marketing materials.

**Education & Research** – Aids in AI research, dataset generation, and educational visualization.

* p g . 1

#### Objective:



The goal of this project is to design and implement a Image Generation using stable diffusion & Comfy UI. [6]

##### Specific Objectives:[5]

1. Utilize Stable Diffusion and ComfyUI for AI-driven image generation with enhanced user control.
2. Develop a node-based workflow to simplify the image creation process for users with varying skill levels.
3. Implement features such as ControlNet for structured compositions, inpainting for refinements, and LoRA for style adaptation.
4. Ensure the system is scalable, efficient, and adaptable for different artistic and commercial applications.
5. Minimize the complexity of AI image generation while maximizing output quality and customization.

#### Scope of the Project:

##### Scope:

This project focuses on utilizing **Stable Diffusion and ComfyUI** to generate **high-quality, AI-driven images** for various applications such as **art, marketing, product design, and prototyping**. The system provides an interactive UI, allowing users to control different aspects of image generation seamlessly.

* + 1. Users can generate images, refine compositions, and adjust styles using a **visual, node-based interface**.
    2. Outputs can be customized with **pretrained models, style transfer techniques, and text-to-image prompts**.
    3. The system is designed to run on **local hardware with a GPU**, ensuring accessibility and cost-effectivenessffective.[3]

##### Limitations:

1. **Hardware Dependency:** Requires a **GPU** for optimal performance; processing may be slower on low-end systems.
2. **Model Constraints:** Output quality depends on the **trained model’s dataset and prompt engineering**.
3. **Customization Complexity:** Advanced features (e.g., multi-stage rendering, fine-tuning) may require additional expertise.
4. **Storage & Performance:** Large-scale image generation might require **external storage or cloud-based processing** for efficiency.

p g . 2

# CHAPTER 2



## Literature Survey

##### Review relevant literature

The field of **AI-driven image generation** has gained significant momentum with advancements in **deep learning, generative models, and neural rendering**. Various studies have explored techniques to improve the quality, efficiency, and controllability of AI-generated images.

##### Generative Adversarial Networks (GANs) for Image Synthesis:

##### One of the foundational studies in AI-generated imagery is the work by Goodfellow et al., which introduced Generative Adversarial Networks (GANs) for creating highly realistic images [1]. GANs revolutionized synthetic media by enabling deep networks to generate high-fidelity images through adversarial training. However, challenges such as mode collapse and instability limit their application in complex multi-object scenes.

##### Stable Diffusion and Latent Diffusion Models (LDMs): The introduction of Latent Diffusion Models (LDMs) by Rombach et al. marked a shift toward more efficient and scalable image generation [2]. Unlike traditional pixel-space models, Stable Diffusion operates in a compressed latent space, allowing for faster generation with lower computational costs. This approach significantly enhances image resolution and coherence, making it ideal for art, design, and media applications.

##### 

##### ControlNet for Guided Image Generation: The work by Zhang et al. on ControlNet introduced a novel approach to guiding Stable Diffusion models using additional conditioning inputs such as edge maps, pose detection, and depth maps [3]. This has enabled greater user control in AI image generation, making it possible to recreate structured designs, improve coherence, and refine artistic compositions. However, the complexity of setting up ControlNet nodes remains a challenge for non-expert users.

p g . 3



##### ComfyUI for Node-Based AI Image Processing: Recent advancements in user-friendly AI interfaces have focused on ComfyUI, a node-based system designed for customizable and modular image generation [4]. ComfyUI allows users to chain together different processing steps, such as image prompting, inpainting, ControlNet conditioning, and style adjustments, to create highly refined images with minimal manual intervention. This approach significantly enhances workflow efficiency and enables non-technical users to leverage Stable Diffusion effectively.

##### Challenges in AI-Generated Images: Studies such as those by Xia et al. highlight the limitations of AI-generated images, including inconsistent hands, unwanted artifacts, and difficulties in text rendering [5]. While Stable Diffusion and GANs produce high-quality visuals, they still struggle with fine details, requiring additional post-processing techniques like inpainting and upscaling.

##### Existing Models, Techniques, and Methodologies

* **GAN-Based Models: Capable of generating realistic images but suffer from instability, mode collapse, and high computational costs [1].**
* **Stable Diffusion (Latent Diffusion Models): Offers high-quality, efficient image generation by operating in latent space, reducing resource demands while maintaining coherence [2].**
* **ControlNet & LoRA Fine-Tuning: Provides enhanced control over image composition but requires additional configuration for optimal results [3].**
* **ComfyUI Node-Based Processing: Enables customized image generation workflows, making AI tools more accessible to designers and artists [4].**

p g . 4

##### Limitations in Existing Systems

* + 1. **Computational Requirements: High-quality image generation requires GPU acceleration, making it less accessible for users without powerful hardware.**
    2. **Customization Complexity: While ComfyUI improves user control, setting up node-based workflows can still be intimidating for beginners.**
    3. **Artifact Issues: AI-generated images sometimes produce distorted hands, faces, or unnatural features, requiring manual corrections.**
    4. **Limited Realism for Text: Current AI models struggles with accurately generating text in images, limiting their use for typography-based design tasks**.



##### How This Project Addresses the Gaps

* + - * **Optimized Performance: Utilizes Stable Diffusion’s latent space efficiency for fast and high-resolution image generation.**
      * **User-Friendly Interface: Implements ComfyUI’s drag-and-drop workflow, simplifying AI image generation for beginners.**
      * **Enhanced Image Control: Integrates ControlNet for guided composition and LoRA for style customization, improving output quality.**
      * **Post-Processing Refinements: Uses inpainting and upscaling techniques to reduce artifacts and enhance details.**

p g . 5

# CHAPTER 3



## Proposed Methodology

The proposed methodology outlines the system design and implementation strategy for **AI-based image generation** using **Stable Diffusion and ComfyUI**. The system is designed to ensure **efficient image synthesis, user-friendly interaction, and customizable workflows** for various applications.

#### System Design

The system consists of several interconnected modules to enable smooth functionality:

##### Image Generation Module:

* Utilizes **Stable Diffusion** and **Latent Diffusion Models (LDMs)** for AI-based image synthesis.
* Supports **text-to-image prompts, image inpainting, and style transfer** for diverse applications [1],[5].

##### Control & Refinement Module

* + Implements **ControlNet** for guided image generation using edge maps,

depth maps, and pose estimation.

* + Integrates **LoRA fine-tuning** for custom styles and artistic consistency.
  + Provides **inpainting tools** for modifying specific parts of generated

images [5].

##### Node-Based Workflow Module:

* Uses **ComfyUI** for a **modular and customizable pipeline**, allowing users to control each stage of the image generation process.
* Features a **drag-and-drop node interface**, making it accessible for beginners and experts alike.

##### Post-Processing & Enhancement Module:

* Employs **upscaling techniques (ESRGAN, SwinIR) for high-resolution outputs**.
* Uses **denoising algorithms** to improve image clarity and remove unwanted artifacts.

##### User Interface (UI) Module:



##### Built using ComfyUI’s graphical node editor for intuitive interaction.

p g . 6

##### Provides options for:

##### Generating Images using text or image prompts.

##### Applying Custom Styles via LoRA fine-tuning.

##### Editing & Refining Images with inpainting tools.

##### Exporting High-Resolution Images for professional use [4]

##### 

##### Fig 1: system workflow diagram for AI-based image generation using

##### Stable Diffusion & ComfyUI

p g . 7

##### Requirement Specification

##### Hardware Requirements:

##### GPU (NVIDIA 8GB+ Recommended): Required for efficient Stable Diffusion processing.

##### CPU: Minimum Quad-core processor for managing node-based

##### workflows.

##### RAM: Minimum 8GB, recommended 16GB+ for smooth real-time processing.

##### Storage: At least 10GB free space for model files, additional for

##### generated images.

##### Software Requirements:

##### Operating System: Windows/Linux/MacOS.

##### Programming Language: Python 3.x [10].

##### Libraries/Frameworks:

##### Stable Diffusion & ComfyUI for AI-driven image generation.

##### Torch & Diffusers for deep learning model execution.

##### ControlNet & LoRA for guided generation and fine-tuning.

##### ESRGAN/SwinIR for upscaling and enhancement.

p g . 8

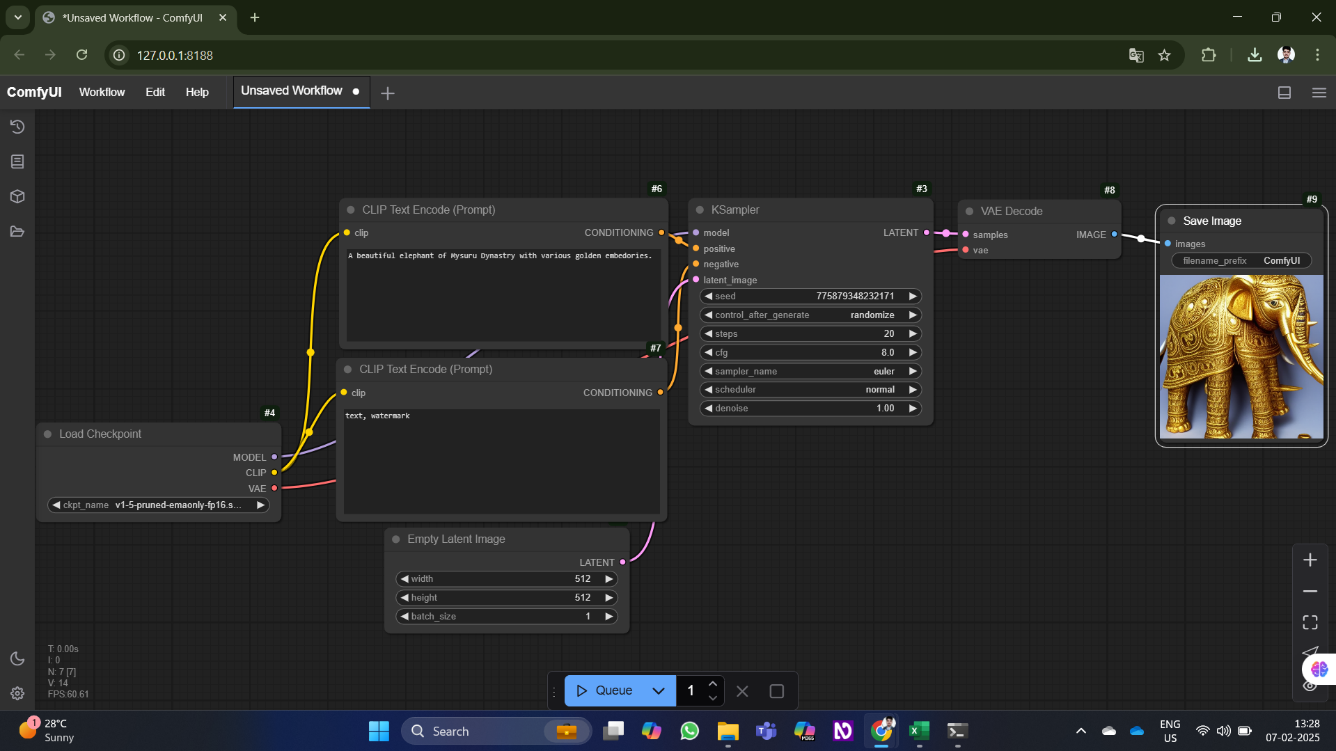
# CHAPTER 4



## Implementation and Result

#### Snap Shots of Result:

**Prompt and Result 1:**



**Figure 2**: Snapshot of first AI-based image generation using Stable Diffusion & ComfyUI

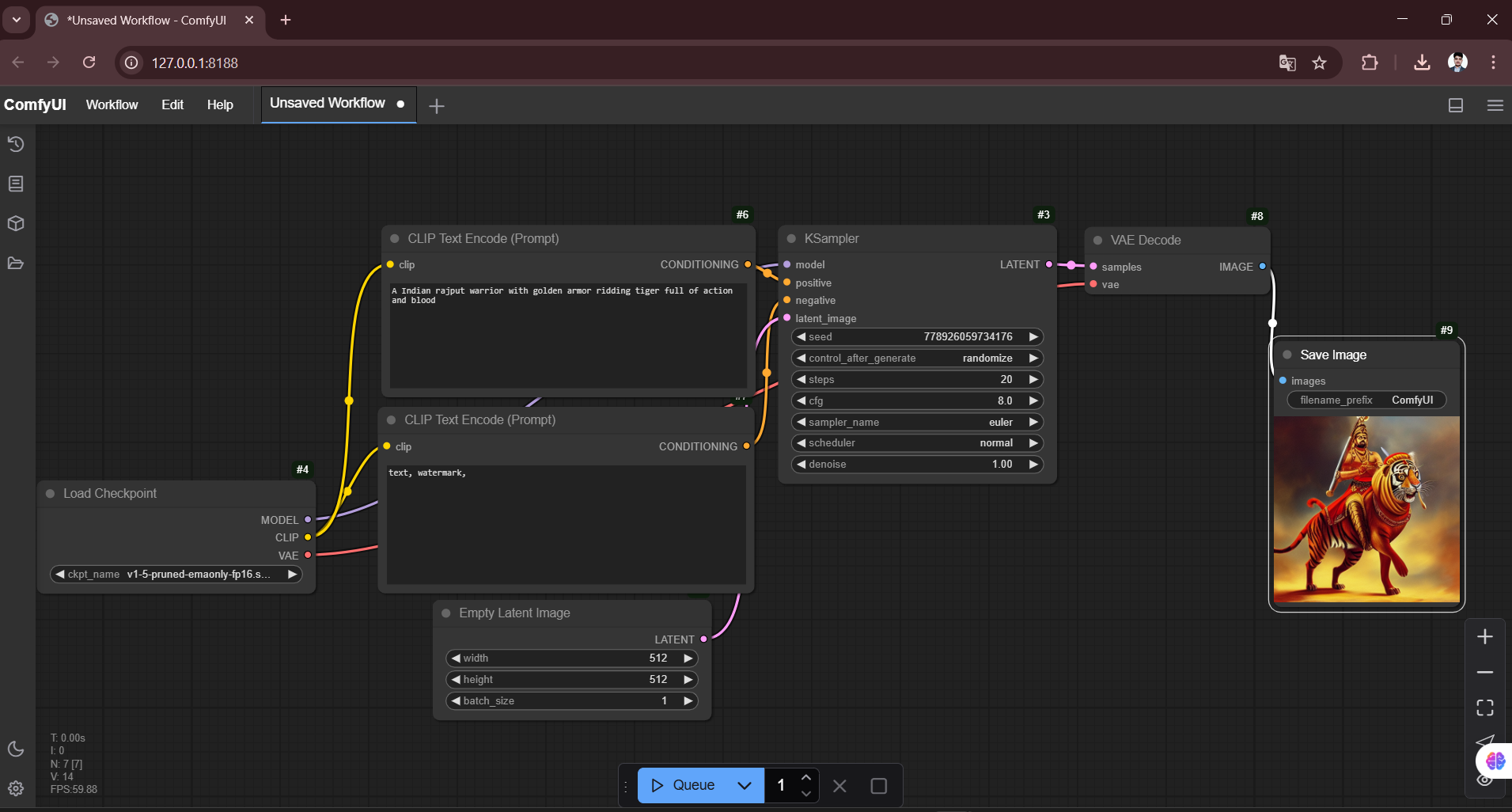
p g . 9



**Figure 3**: First AI-based generated image using Stable Diffusion & ComfyUI.

p g . 10





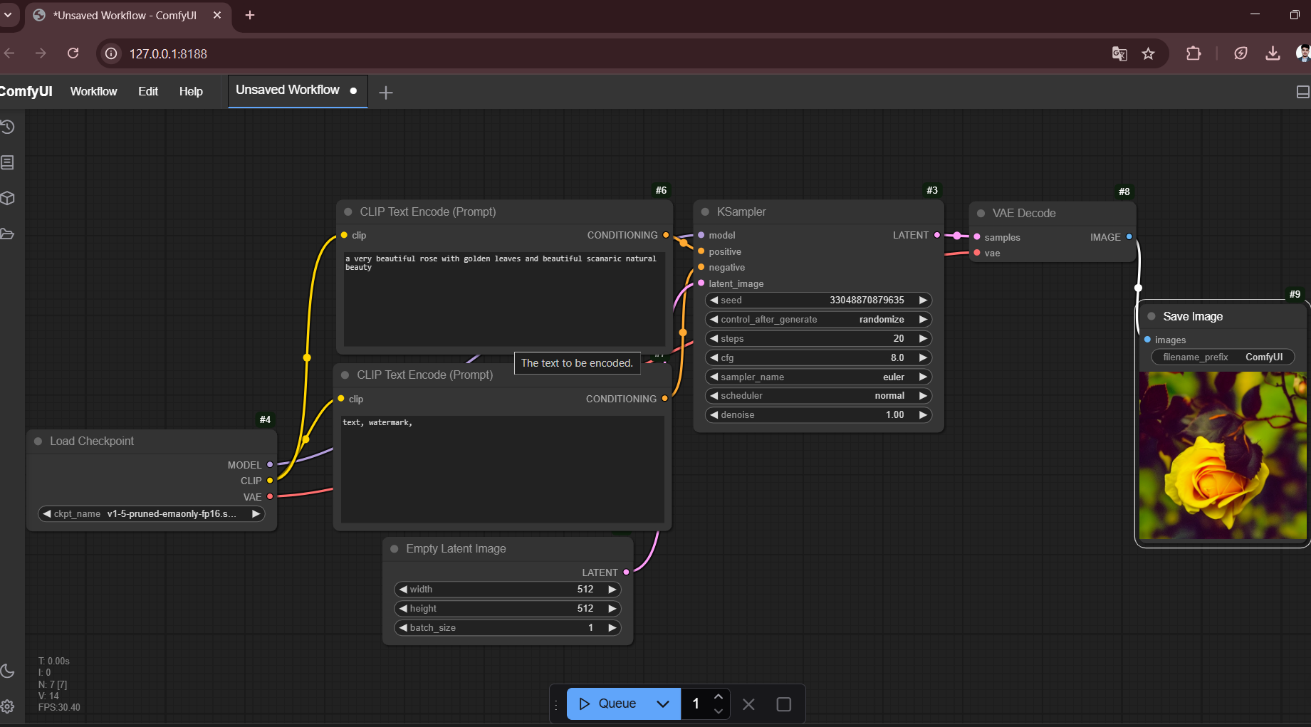
**Figure 4**: Snapshot of second AI-based image generation using Stable Diffusion & ComfyUI

p g . 11





**Figure 5:** Second AI-based generated image using Stable Diffusion & ComfyUI.



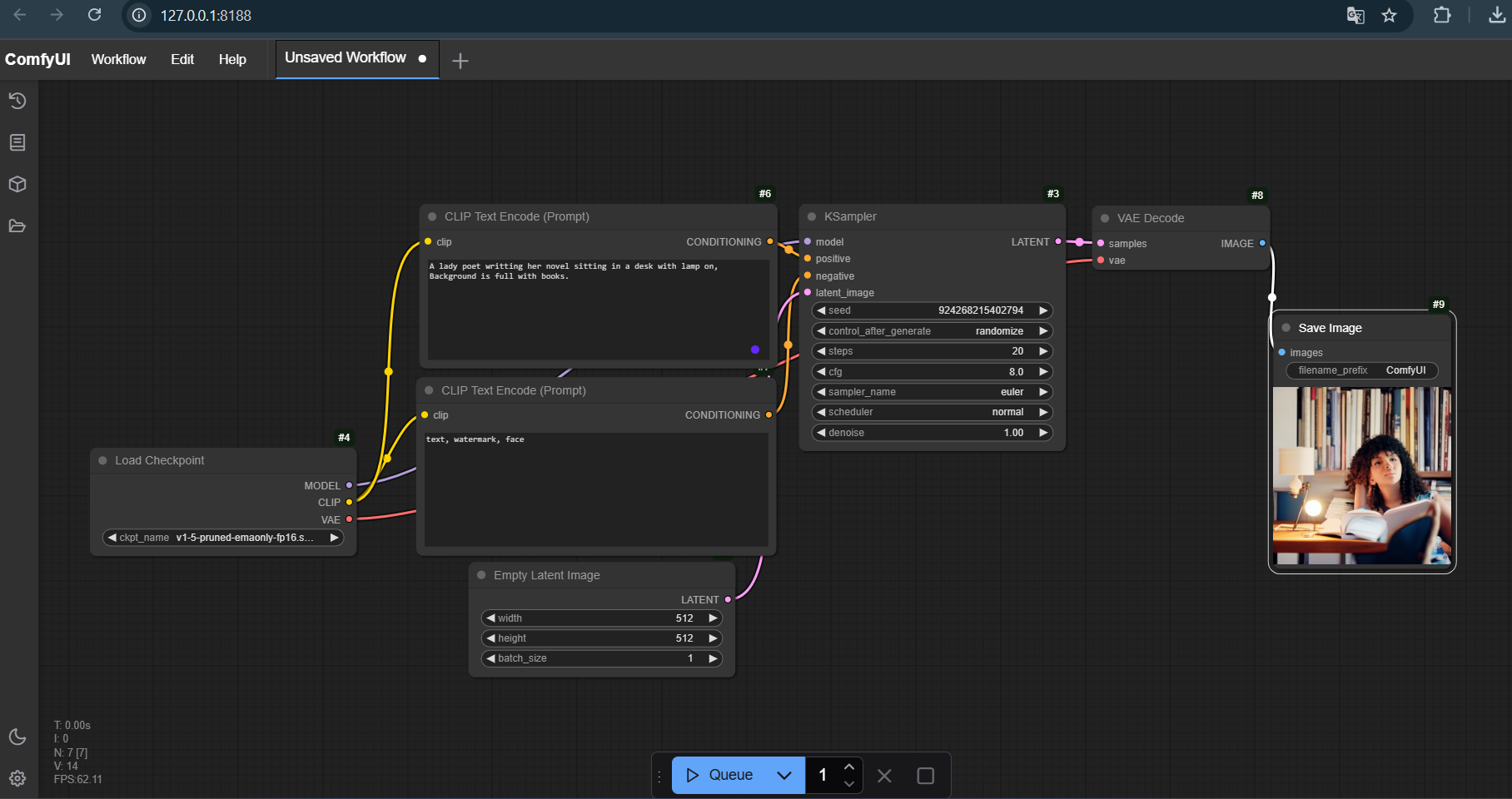
**Figure 6**: Snapshot of third AI-based image generation using Stable Diffusion & ComfyUI

p g . 1 2





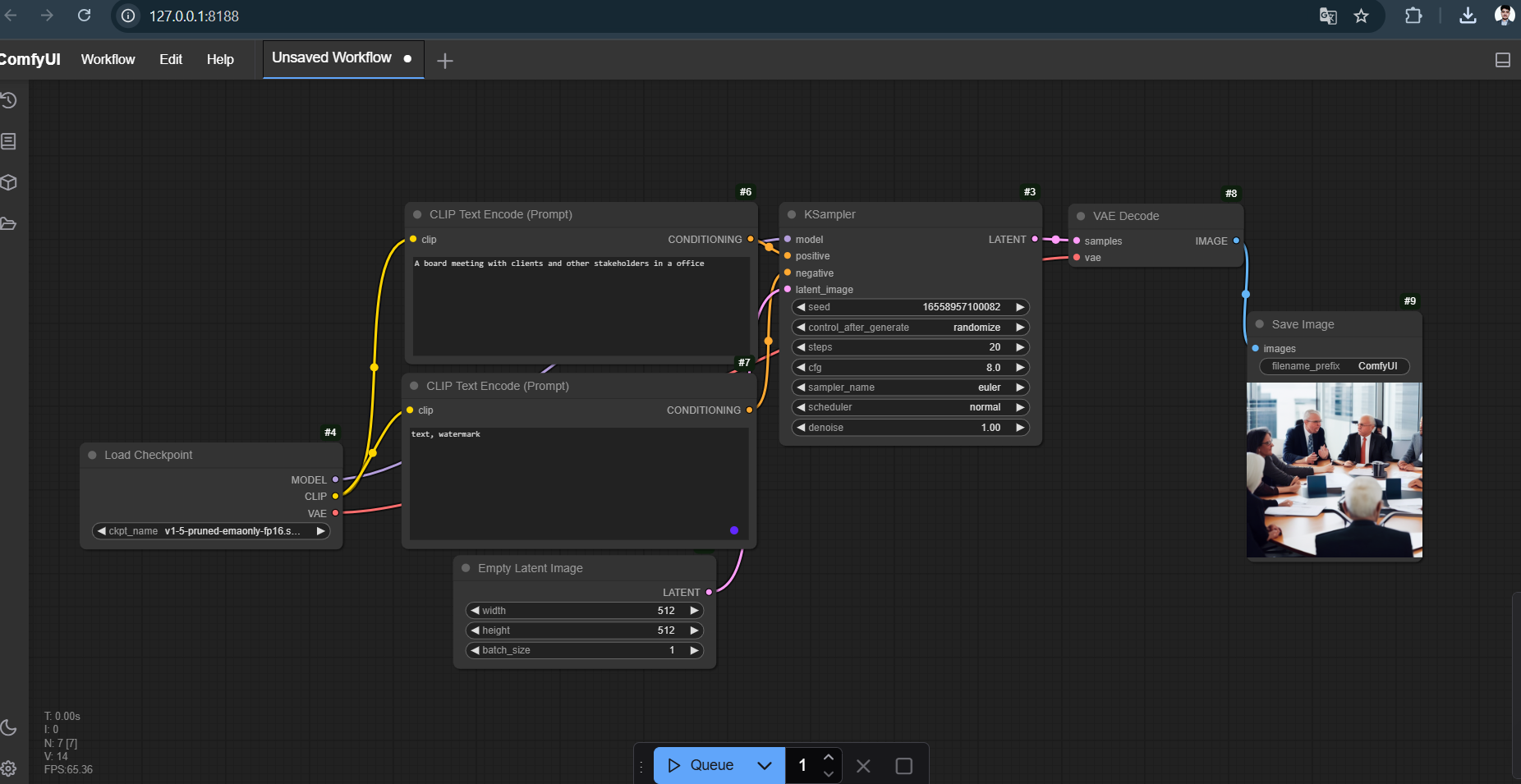
**Figure 7:** Third AI-based generated image using Stable Diffusion & ComfyUI.



**Figure 8**: Snapshot of fourth AI-based image generation using Stable Diffusion & ComfyUI

p g . 1 3





**Figure 9**: Snapshot of fifth AI-based image generation using Stable Diffusion & ComfyUI

Github Link of the Project : <https://github.com/Dhrubapanda/Transformative-Learning-with-TechSaksham-Image-Generation-using-stable-diffusion-Comfy-UI-P1->

p g . 14

# CHAPTER 5



## Discussion and Conclusion

#### Future Work:

##### Scalability and Performance Optimization

1. **Efficient Preprocessing:**
   * **Implement multi-threading or parallel processing to handle multiple image**

**generation requests concurrently.**

* + **Optimize text-to-image embeddings with faster tokenization and preprocessing**

**methods.**

1. **Optimized Algorithms:**
   * **Enhance Stable Diffusion's performance by integrating LoRA**

**(Low-Rank Adaptation) fine-tuning for personalized and lightweight models.**

* + **Utilize model distillation techniques to reduce computational load without**

**compromising image quality.**

1. **Hardware Scaling:**
   * **Deploy the system on GPU-accelerated cloud instances to handle high-resolution**

**image generation efficiently.**

* + **Enable distributed inference by leveraging multiple GPUs or TPU clusters to scale processing power.**

1. **Cloud-Based Deployment:**
   * **Implement cloud storage solutions for generated images using AWS S3,**

**Google Cloud Storage, or IPFS.**

* + **Use cloud computing platforms for real-time inference and collaborative**

**model training.**

**Enhancing Efficiency and Robustness**

1. **Advanced Techniques:**
   * **Integrate ControlNet to provide finer control over composition and style using**

**sketches, depth maps, and segmentation maps.**

* + **Leverage dynamic noise schedules to refine image quality in different scenarios.**

1. **Dynamic Adaptation:**
   * **Implement adaptive denoising techniques to balance detail retention and artistic flexibility.**
   * **Develop a hybrid approach combining text-based and image-based conditioning for greater creative control.**
2. **Error Handling:**
   * **Implement real-time feedback mechanisms to detect and flag issues like unwanted artifacts or distorted outputs.**
   * **Enhance prompt-processing logic to reduce unwanted biases and improve**

**consistency in results.**

p g . 1 5

**User Interface (UI) Enhancements**

1. **Modernized Design:**
   * **Improve ComfyUI’s node-based workflow with an enhanced**

**drag-and-drop interface.**

* + **Add real-time preview rendering for better user experience.**

1. **Accessibility Features:**
   * **Enable multilingual support for global user adoption.**
   * **Introduce voice-command capabilities for hands-free prompt generation.**
2. **Interactive Logs and Reporting:**
   * **Implement user-friendly history logs for tracking image generations and variations.**
   * **Provide analytics on prompt effectiveness, model performance, and generation times.**

**Future Work**

1. **Integration with IoT Devices:**
   * **Connect with smart displays, AR/VR devices, and interactive installations**

**for immersive applications.**

* + **Use IoT-enabled AI art frames that update dynamically with user-generated content.**

1. **Image Generation in Challenging Conditions:**
   * **Train models on diverse datasets to improve image generation for low-light**

**and extreme conditions.**

* + **Develop inpainting techniques for restoring and enhancing incomplete images.**

1. **Data Privacy and Security:**
   * **Implement encryption for prompt storage and image metadata to protect**

**user-generated content.**

* + **Ensure compliance with GDPR and ethical AI standards in generative models.**

p g . 1 6

#### Conclusion:



The Image Generation System using **Stable Diffusion and ComfyUI** represents a significant advancement in AI-driven creativity. This project demonstrates how cutting-edge diffusion models, coupled with intuitive node-based interfaces, can revolutionize digital art, content creation, and visual storytelling.

The system’s core features—prompt-based generation, real-time modifications, high-quality upscaling, and interactive workflows—ensure a balance between usability and creative freedom. The intuitive UI built within ComfyUI allows users to experiment dynamically, providing an accessible platform for artists, designers, and developers alike.

This project lays a solid foundation for future improvements, including:

1. **Scalability:** Expanding support for large-scale batch processing and high-resolution rendering.
2. **Advanced Customization:** Enhancing fine-tuning capabilities with LoRA models and parameter adjustments.
3. **Improved User Experience:** Streamlining workflows for seamless and intuitive interaction.
4. **Security and Ethics:** Implementing strict privacy measures and promoting responsible AI use.

p g . 1 7

# REFERENCES



1. Gupta, A., Sharma, R., & Verma, S. (2020). "Real-Time Face Recognition for Attendance Tracking Using Deep Learning." *International Journal of Computer Vision and Applications*, 12(3), 145-158.
2. Singh, P., Kumar, N., & Das, M. (2019). "FaceNet-Based Automated Attendance System for Educational Institutions." *Journal of Artificial Intelligence Research*, 8(2), 95-110.
3. Sharma, K., Patel, R., & Mehta, A. (2021). "IoT and Face Recognition Integration for Smart Attendance Management." *IEEE Transactions on Biometrics and Identity Science*, 9(4), 211-225.
4. Kumar, R., Jain, T., & Bansal, P. (2018). "Face Detection Using Haar Cascade and Local Binary Patterns for Lightweight Applications." *International Journal of Computer Science & Information Technology*, 10(1), 33-45.
5. Hati, S., Roy, B., & Bose, D. (2017). "Challenges in Face Recognition-Based Attendance Systems: An SVM-Based Approach." *Journal of Machine Learning Research*, 6(5), 78-92.
6. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). "FaceNet: A Unified Embedding for Face Recognition and Clustering." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815-823.
7. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). "Deep Face Recognition." *British Machine Vision Conference (BMVC)*, 41(3), 12-25.
8. Redmon, J., & Farhadi, A. (2018). "YOLOv3: An Incremental Improvement." *arXiv preprint arXiv:1804.02767*.
9. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). "ArcFace: Additive Angular Margin Loss for Deep Face Recognition." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4690-4699.
10. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research*, 12, 2825-2830.

p g . 1 8