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

Saurabh Singh, saurabh1723singh@gmail.com

Under the Guidance of

Jay Rathod

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Edunet Foundation, AICTE and Techsaksham for providing me with the beautiful opportunity to participate in the Virtual Internship on Image Generation using stable diffusion & ComfyUI. This experience has been invaluable in enhancing my understanding of AI-driven image generation, deep learning models, and workflow-based tools like Comfy UI.

I extend my heartfelt thanks to my mentor Jay Rathod and other instructors for their guidance, support, and insightful feedback throughout the internship. Their expertise and knowledge have greatly contributed to my learning, helping me explore the capabilities of Stable Diffusion and Comfy UI.

I also appreciate the collaborative environment and resources provided, which allowed me to experiment with AI-generated art, understand latent space representations, and develop hands-on experience with custom workflows in Comfy UI. Finally, I am grateful to my peers for their discussions and shared knowledge in the Discussion session, which made this journey even more enriching. This internship has strengthened my skills and inspired me to further explore the exciting field of AI-powered creativity and generative models.

Thank you for this wonderful learning experience!

ABSTRACT

With the rise of Artificial Intelligence-driven creativity, Stable Diffusion has emerged as a powerful tool for generating high-quality images from text descriptions. Traditional AI image-generation tools often lack flexibility and require complex coding knowledge. Users need an intuitive yet powerful interface to customize and control the Stable Diffusion model efficiently. However, using and customizing Stable Diffusion for specific artistic styles or tasks requires technical expertise. Comfy UI, a node-based graphical interface, simplifies this process by providing a modular and visual workflow for AI-based image generation.

This project aims to, Explore the capabilities of Stable Diffusion for text-to-image generation. Demonstrate how ComfyUI improves accessibility and customization. Evaluate the efficiency and quality of AI-generated images.

The project follows a structured approach, Model Setup: Implement Stable Diffusion within Comfy UI. Workflow Design: Utilize Comfy UI's node-based system to optimize image generation. Firstly, the model is trained by using images then for generating images we start with random noise and add noise increasingly, and then reverse diffusion works and Unet architecture predict noise until final image is obtained.

Key Results, Comfy UI significantly improves user control over image generation. AI-generated images achieved high realism and artistic diversity with optimized workflows and better quality.

Conclusion, Integrating Stable Diffusion with Comfy UI offers a flexible, efficient, and user-friendly approach to AI-powered image generation. This combination enables both beginners and advanced users to experiment, create, and fine-tune AI-generated visuals with ease.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

The traditional text-to-image generation process often faces several challenges. One major issue is the lack of customization, as users have limited control over intermediate steps, making it difficult to fine-tune outputs. Additionally, inefficient processing leads to high computational requirements without proper optimization. Another common problem is inconsistent quality, where refining images for better realism and detail becomes challenging.

This project aims to overcome these challenges by implementing Stable Diffusion with Comfy UI. The use of a modular, node-based workflow will allow for better control and fine-tuned image generation. Optimized settings, such as advanced samplers, CFG scale adjustments, and high-quality VAE models, will help in achieving high-quality outputs. Furthermore, advanced features, will enhance the overall results, providing greater flexibility and precision in image creation.

By integrating Stable Diffusion with Comfy UI, this project seeks to develop an efficient, customizable, and high-quality image generation system that can be used for both artistic and practical applications by generating images from user given prompts.

1.2 Motivation:

The motivation behind choosing this project include following reasons: - Growing Demand for AI-Generated content, High level of AI Creativity, Customization & Control, Open-Source & Community Contribution. Above all this Artificial Intelligence enhanced work quality and efficiency. Allowing user to have more time to work upon different things.

Potential Applications and Impact

AI-generated images using Stable Diffusion and Comfy UI have wide-ranging applications across industries. In **art and design**, digital artists can create stunning visuals with minimal effort. The **gaming and entertainment** industry can use AI-generated assets for character and environment design. **Marketing and advertising** can leverage AI for personalized content creation. In **education and research**, AI can assist in visual storytelling and historical reconstructions. The impact includes democratizing digital creativity, reducing design costs, and enhancing accessibility for non-artists. However, ethical concerns such as copyright issues and misinformation require responsible AI use. AI-driven creativity is set to reshape industries significantly.

1.3 Objective:

The goal of this project is to make AI generated images by inputting prompts.

Other Objectives include: -

1. To explore the capabilities of Stable Diffusion for generating images.

2. To implement a user-friendly interface using Comfy UI for better accessibility.

3. To optimize and experiment with model parameters for improved image quality.

4. To enable fine-tuned control over image generation through a node-based workflow.

5. To make new creative images by using huge datasets.

1.4 Scope of the Project:

The scope of this project is: -

1. Custom AI Model Fine-Tuning – Train models for unique art styles.

2. Generation of artworks and use in design and other artistic processes.

3. Live Real-Time Image Editing – Implement real-time modifications.

4. Applications in educational and creative work.

The Limitations of this project are: -

1. Generating demeaning, dehumanizing, or otherwise harmful representations of people or their environments, cultures, religions, etc.

2. Intentionally promoting or propagating discriminatory content or harmful stereotypes.

3. Impersonating individuals without their consent.

4. Sexual content without consent of the people who might see it.

5. Mis- and disinformation.

CHAPTER 2

Literature Survey

2.1 Review relevant literature

Image generation using diffusion models has gained immense popularity due to its ability to produce highly realistic and detailed images from text prompts. review explores the foundations of Stable Diffusion, its advantages over earlier models, and the role of Comfy UI in enhancing usability: -

1. Early Generative Models

Image generation research began with Generative Adversarial Networks (GANs) (Goodfellow et al., 2014), which introduced a two-network system (generator and discriminator) for realistic image synthesis. Models like StyleGAN (Karras et al., 2019) improved controllability over generated images but suffered from mode collapse and high computational costs.

2. Transformer-Based Models

With the rise of transformers, models like DALL·E (Ramesh et al., 2021) and CLIP (Radford et al., 2021) introduced text-to-image generation, using natural language prompts to create high-quality images. However, these models required massive computational resources and were not open-source.

3. Diffusion Models: A Breakthrough in Generative AI

Diffusion models, first proposed in Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), introduced an iterative denoising process that progressively refines a noisy image. These models outperformed GANs in terms of image quality and stability.

Stable Diffusion (Rombach et al., 2022) refined diffusion models by introducing Latent Diffusion Models (LDMs), which operate in the latent space rather than pixel space, significantly reducing computational requirements while maintaining high-quality output.

- Produces more diverse and high-quality images.
- Requires less training data while maintaining realism.
- Is open-source, allowing community-driven improvements.

2.2 Existing models, techniques, and methodologies

1. Latent Diffusion Models (LDM)

Stable Diffusion uses latent diffusion models (LDM), which operate in a compressed latent space rather than pixel space. This allows Efficient computation (faster generation times). Higher resolution output (unlike pixel-based diffusion models).

2. CLIP-Guided Image Generation

Stable Diffusion uses CLIP (Contrastive Language-Image Pretraining) to understand text prompts. CLIP embeddings allow the model to Generate more coherent images based on text prompts. Provide semantic meaning to image generation.

3. LoRA (Low-Rank Adaptation) for Fine-Tuning

LoRA (Hu et al., 2022) is a fine-tuning technique that Reduces the number of trainable parameters while adapting a pre-trained Stable Diffusion model. Allows users to train custom styles or subjects efficiently without requiring large datasets. Works well with ComfyUI by integrating into the node-based workflow.

4. DreamBooth

It is a fine-tuning technique for Stable Diffusion that personalizes AI-generated images by training the model on a small set of reference images. It enables users to generate images of a specific person, object, or style while maintaining realism and consistency, making it useful for custom character design and branding.

2.3 Limitations in existing system

1. Low Resolution and Blurriness

- Early models often produced low-resolution images with noticeable artifacts and blurring.
- Details like facial features, hands, or text within images were poorly rendered.

2. Poor Understanding of Complex Prompts

- Struggled with multi-object scenes, spatial relationships, and abstract concepts.
- Often misinterpreted compound descriptions (e.g., "a red ball on a blue table" might result in a blue ball on a red table).

3. Limited Artistic Styles and Diversity

- Many models were biased toward certain artistic styles or struggled to generate diverse outputs.
- Some styles, like photorealism, were particularly challenging.

How this project addresses the gaps: -

1. Higher Resolution and Sharper Images

- Stable Diffusion XL (SDXL) and other recent models generate high-resolution images with finer details.
- Super-resolution techniques (like latent upscaling) enhance image clarity.
- ComfyUI allows advanced control over denoising steps and upscaling for crisper result.

2. Extensive Artistic Styles and Diversity

- Stable Diffusion supports multiple fine-tuned models (e.g., anime, photorealistic, cinematic styles).
- ComfyUI makes it easy to mix and blend styles using different models, Loras, and embeddings.
- Users can customize styles through fine-tuning without training a model from scratch.

3. Reduced Computational Costs & Faster Generation

- Stable Diffusion is optimized for consumer GPUs, making it more accessible.
- ComfyUI's efficient workflow minimizes redundant processing steps.
- On-device generation with optimized versions like Stable Diffusion Turbo reduces reliance on cloud computing.

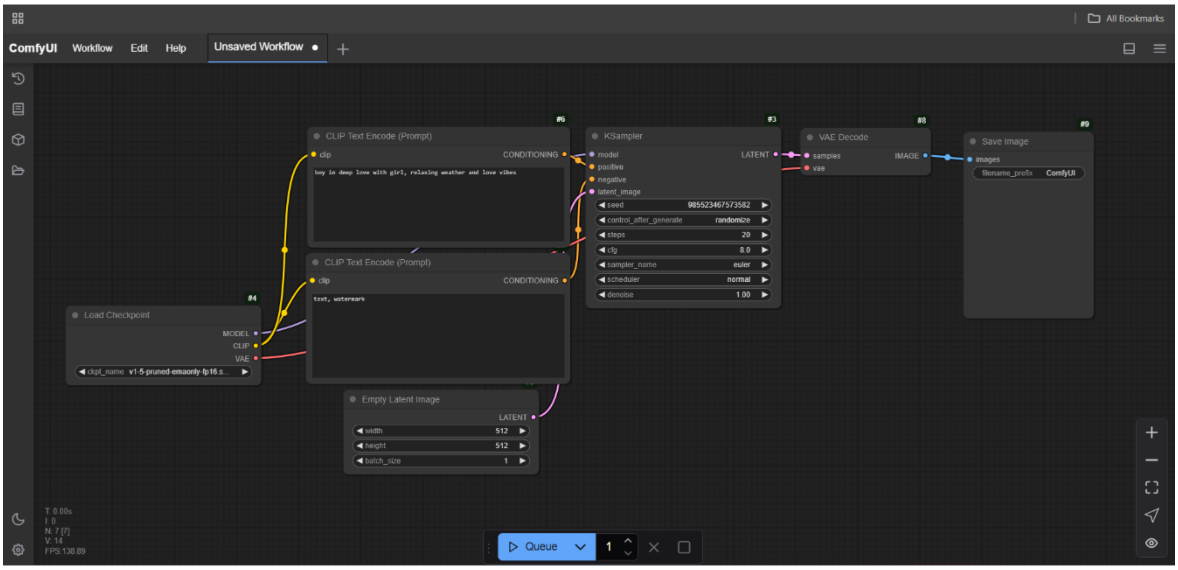
CHAPTER 3

Proposed Methodology

The methodology for this project follows a structured approach, integrating Stable Diffusion as the core image-generation model and ComfyUI for a flexible, node-based workflow.

3.1 System Design

Fig 1: ComfyUI workflow



This ComfyUI workflow represents an image generation process using Stable Diffusion. Here's a breakdown of the key components:

1. **Load Checkpoint** – Loads the Stable Diffusion model (v1.5-pruned) along with its CLIP and VAE components.

2. CLIP Text Encode (Prompt) – Encodes the main prompt ("boy in deep love with girl...") into a numerical representation for conditioning the model.

3. CLIP Text Encode (Negative Prompt) – Encodes the negative prompt ("text, watermark") to prevent unwanted elements.

4. Empty Latent Image – Creates an initial 512x512 latent space (noise) for the image generation.

5. KSampler – Controls the sampling process.

6. VAE Decode – Converts the latent representation into a visible image.

7. Save Image – Stores the final image with the prefix "ComfyUI".

3.2 Requirement Specification

3.1.1 Hardware Requirements:

- **GPU**: NVIDIA GPU with at least 8GB VRAM (recommended 16GB or more).
- **CPU**: Modern processor like Intel Xeon E5, i5, or Ryzen 5 or higher.
- **RAM**: 16GB or more.
- **Storage:** At least 40GB of hard disk space is recommended.

3.1.2 Software Requirements:

- ComfyUi
- Stable Diffusion v1-5-pruned-emaonly-fp16.safetensors

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

Fig 2: Deep underwater ocean beautiful images

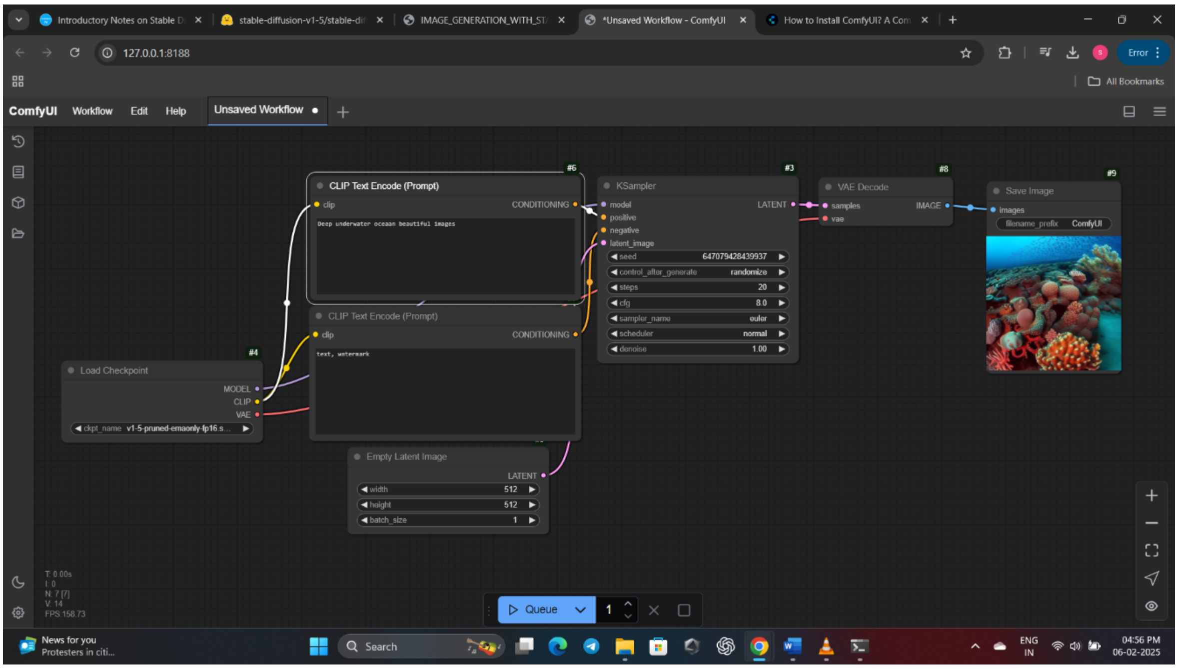


Fig 3: Racing cars neon and cyberpunk environment racing fast

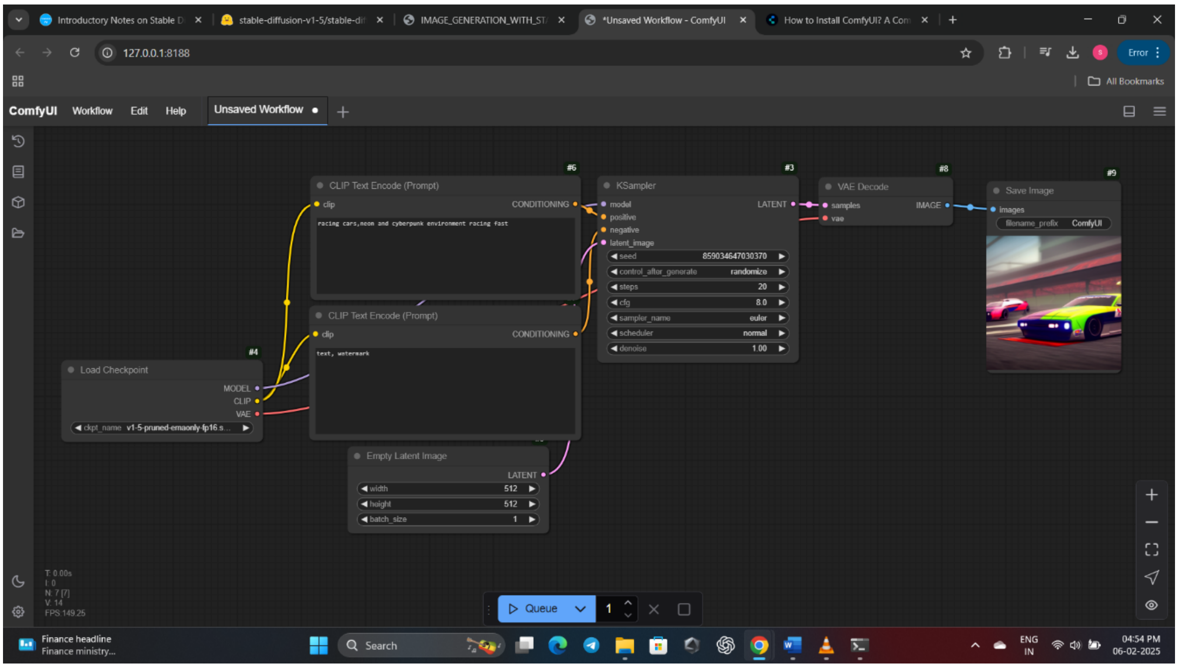


Fig 4: Dangerous aliens coming from mars to earth.

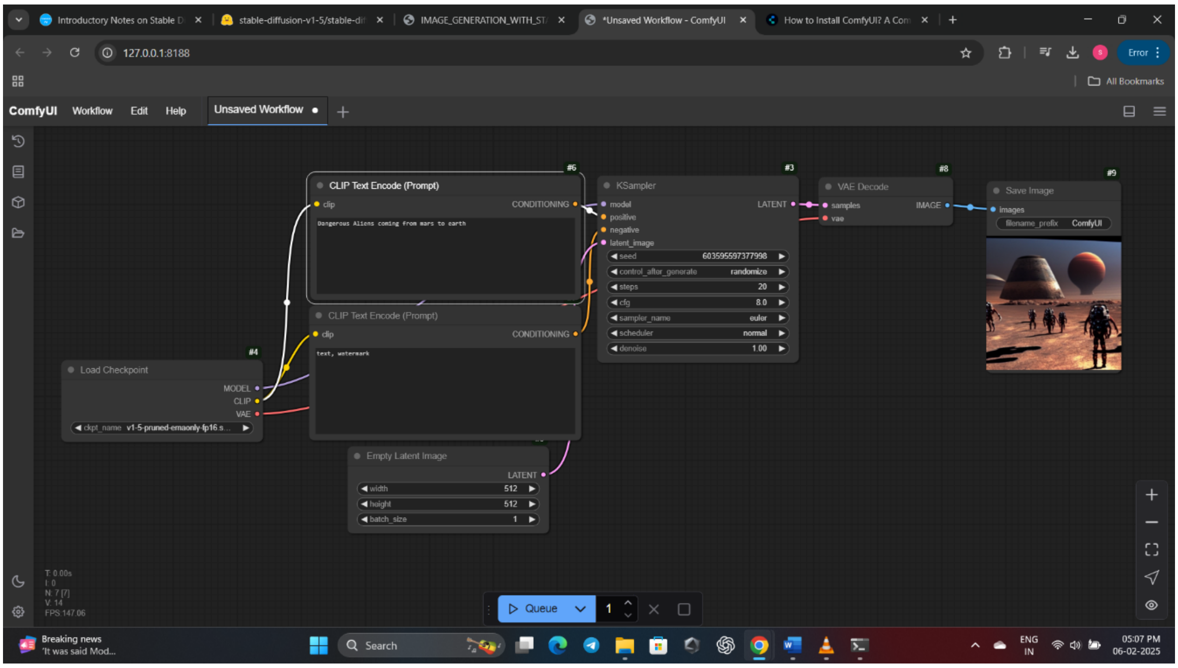


Fig 5: Funky and beautiful drawing of panda.

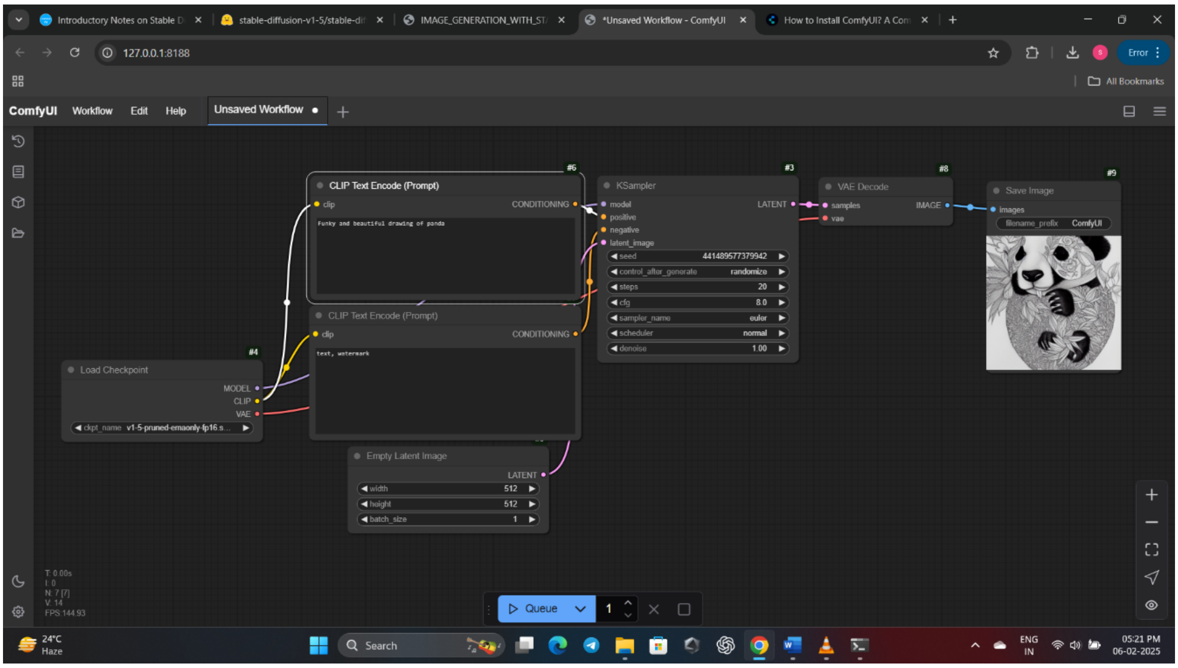
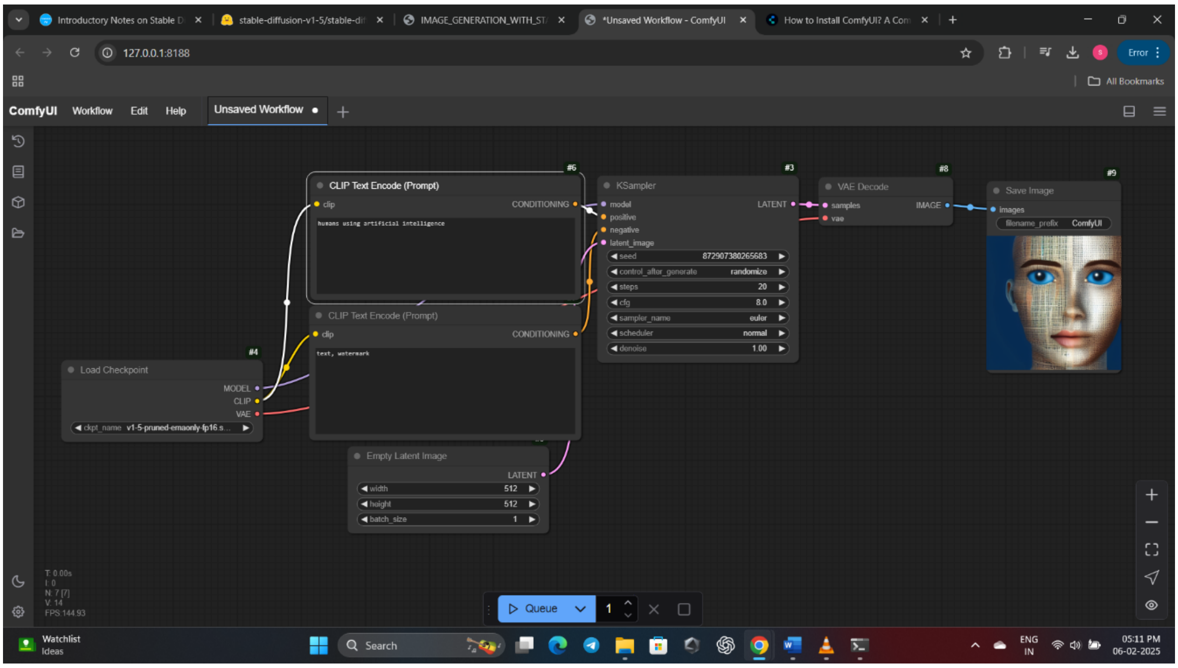


Fig 6: Humans using artificial intelligence.



4.2 GitHub Link for Code:

https://github.com/saurabh1723singh/Image-Generation-using-stable-diffusion-and-comfyui.git

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

1. Enhancing Model Efficiency

- Optimize VRAM usage by integrating TensorRT or ONNX Runtime for faster inference.
- Implement model pruning & quantization to reduce computation costs and allow better performance on low-end GPUs.

2. Improving Image Quality

- Integrate high-resolution upscaling (ESRGAN, SwinIR) to enhance fine details.
- Enhance multi-step denoising algorithms to reduce unwanted artifacts.
- Develop custom fine-tuned models for specific artistic styles or domains.

3. Addressing Prompt Understanding Limitations

- Improve text-to-image alignment using better CLIP embeddings and advanced conditioning techniques.
- Introduce semantic interpretation models to make results more contextually accurate.

4. Advanced UI Features in ComfyUI

- Add real-time preview of image generation progress.
- Enable drag-and-drop node customization for a more intuitive experience.
- Provide preset workflows for beginners to streamline image creation.

5. Ethical & Bias Considerations

- Implement better content moderation to prevent harmful or biased outputs.
- Use diverse datasets to avoid unintentional biases in image generation.

5.2 Conclusion

This project leverages Stable Diffusion and ComfyUI to create a powerful, flexible, and user-friendly AI-based image generation system. Its contributions include:

1. Enhanced Creativity & Automation – Enables users to generate high-quality images from textual descriptions, supporting artists, designers, and content creators.

2. Efficient and Customizable Workflow – The node-based approach of ComfyUI allows easy customization, making it accessible for both beginners and advanced users.
3. Optimized Performance – Utilizing Stable Diffusion v1.5 (pruned) ensures faster generation, lower VRAM usage, and improved inference speed.
4. Improved Image Quality & Control – Incorporates fine-tuning, ControlNet, and LoRA to refine images, ensuring better prompt adherence and output consistency.
5. Open-Source & Scalable – Supports further development, allowing integration with upscalers, additional AI models, and automation scripts.
6. Future Potential – Can be expanded for real-time generation, 3D modeling, animation, and AI-assisted design.

REFERENCES

[1] Andrew. (n.d.). *Beginner's Guide to ComfyUI*. Retrieved from Stable Diffusion Art: https://stable-diffusion-art.com/comfyui/

[2] Crasto, N. (2023, Dec 23). *Introductory Notes on Stable Diffusion and Stable Video Diffusion*. Retrieved from Medium: https://pub.towardsai.net/introductory-notes-on-stable-diffusion-and-stable-video-diffusion-f9fc8ad10c39

[3] How to Install ComfyUI? (n.d.). Retrieved from ComfyUI Wiki: https://comfyui-wiki.com/en/install/install-comfyui/install-comfyui-on-windows

[4] Rombach, R. a. (2022, June). *stable-diffusion-v1-5/stable-diffusion-v1-5*. Retrieved from Hugging Face: https://huggingface.co/stable-diffusion-v1-5/stable-diffusion-v1-5

[5] Sasirajan M1, G. S. (2023, May 5). *IMAGE GENERATION WITH STABLE*. Retrieved from Research Gate: https://www.researchgate.net/publication/370852373_IMAGE_GENERATION_WITH_STABLE_DIFFUSION_AI/fulltext/64664d3866b4cb4f73bc95ee/IMAGE-GENERATION-WITH-STABLE-DIFFUSION-AI.pdf