Project Title: AI Image Generator

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

Perumalla Chandra Sekhar,
Email:chandrasekharperumaalla@gmail.com

Under the Guidance of

TechSaksham Coordinator

ACKNOWLEDGEMENT

We would like to take this opportunity to express our heartfelt gratitude to all individuals who contributed directly or indirectly to the successful completion of this project.

First and foremost, I extend my sincere appreciation to my mentor, the **TechSaksham Coordinator**, for their invaluable guidance, unwavering support, and constant motivation throughout this journey. Their insights and expertise have played a crucial role in shaping the direction and execution of this work.

I am also profoundly grateful to **AICTE and TechSaksham** for organizing this internship, which provided me with an incredible opportunity to explore and enhance my skills in **AI-driven image generation**. This experience has been instrumental in broadening my knowledge and fostering my passion for artificial intelligence.

Lastly, I would like to express my deep appreciation to my **peers, colleagues, and everyone who contributed their time, effort, and encouragement** to make this project a success. Their collaboration and support have been invaluable in overcoming challenges and achieving our goals.

ABSTRACT

Artificial Intelligence (AI) has revolutionized creative applications, enabling automated image generation from textual descriptions. This project, **AI Image Generator**, leverages **Stable Diffusion** and **ComfyUI** to generate high-quality images based on user-inputted prompts. The primary objective is to provide an intuitive and efficient AI-powered tool for creative professionals, artists, and researchers to visualize concepts seamlessly.

Problem Statement

Manual image creation requires significant time and skill, while existing AI tools often lack customization and real-time interactivity. This project addresses these challenges by integrating **Stable Diffusion** with a user-friendly interface, allowing users to generate images efficiently based on textual prompts.

Objectives

1. **Develop a user-friendly interface** using **HTML, CSS, and JavaScript** to allow prompt-based image generation.
2. **Implement a robust backend** using **Python and ComfyUI** to process user inputs and generate images with Stable Diffusion.
3. **Optimize the workflow** for seamless interaction, ensuring efficient prompt-to-image conversion.

Methodology

- **Frontend:** A simple and interactive UI featuring a **prompt input field**, a **default prompt ("Sun rising over the seashore")**, and a **dark mode toggle** for accessibility.

- **Backend:** A Python-based **ComfyUI structure** processes user requests, executing Stable Diffusion's step-by-step image generation.
- **Processing Workflow:**
    1. User enters a prompt in the frontend.
    2. The request is sent to the backend.
    3. Backend processes the image step-by-step using Stable Diffusion.
    4. Generated image is returned and displayed.

Key Results

- Successfully developed a **functional AI image generator**.
- Demonstrated **real-time AI-based creativity** through Stable Diffusion.
- Processing time is **7-10 minutes**, influenced by computational complexity.

Conclusion

This project highlights the **potential of AI in creative fields**, offering an automated and interactive solution for **image generation from textual descriptions**. Future improvements may include **faster processing, enhanced image resolution, and additional customization options**.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1Problem Statement:

AI-based image generation has revolutionized the field of digital art and design, allowing users to create stunning visuals from textual descriptions. However, existing tools often lack **user-friendly interfaces** and require technical expertise. This project bridges that gap by developing a **simple, intuitive AI Image Generator** using **Stable Diffusion and ComfyUI**.

1.2Motivation:

The motivation behind this project is to enable **easy and efficient image generation** for users who may not have prior experience in AI or programming. By leveraging **Stable Diffusion**, this project aims to make AI-driven image creation accessible to everyone, from digital artists to content creators.

1.3Objective:

- Develop a **frontend interface** that allows users to enter prompts.
- Implement a **backend pipeline** using **ComfyUI** to process prompts and generate images.
- Ensure a **smooth user experience** with a clear workflow.
- Optimize processing steps for **efficient image generation**.

1.4Scope of the Project:

- **In-scope**:
    - AI-powered image generation using **Stable Diffusion.**
    - User-friendly UI for input and output display.
    - Step-by-step processing in ComfyUI.

- **Out of Scope:**
    - Real-time or instant image generation (processing time is 7-10 minutes).
    - Advanced editing tools within the interface.

CHAPTER 2

Literature Survey

2.1 Introduction

Artificial Intelligence (AI) has revolutionized the creative industry by enabling machines to generate realistic images from textual descriptions. Various AI-driven approaches such as **Generative Adversarial Networks (GANs)**, **Variational Autoencoders (VAEs)**, and **Diffusion Models** have significantly improved the quality of AI-generated images.

Stable Diffusion, a **latent diffusion model**, has gained popularity for its ability to generate **high-quality images** with detailed textures and realistic aesthetics. Comfy UI, a **modular interface**, enhances the user experience by structuring the image generation process into steps, allowing for greater flexibility and control.

2.2 Related Work

Several research papers and AI models have contributed to the advancements in image generation:

- **DALL·E 2** (OpenAI): Uses transformer-based architectures to generate creative and detailed images from text.
- **Imagen** (Google Research): Employs large-scale diffusion models for state-of-the-art image synthesis.
- **Stable Diffusion** (CompVis, LMU): An open-source alternative that balances computational efficiency and image quality.
- **Comfy UI**: Provides a visual node-based workflow to manage diffusion model pipelines effectively.

2.3 Limitations of Existing Approaches

Despite the advancements in AI-based image generation, there are several challenges:

- **High Computational Costs: Generating high-resolution images requires powerful GPUs.**
- **Processing Time: Diffusion-based models often require multiple steps, leading to long generation times.**
- **Fine-tuning Challenges: Customizing AI models for specific artistic styles requires expertise.**

2.4 How Our Project Addresses These Limitations

Our AI Image Generator project improves upon these limitations by:

- **Using ComfyUI to Optimize Workflow: Step-by-step image generation ensures structured processing.**
- **Providing a Simple and Accessible UI: No prior AI expertise is needed for users to generate images.**
- **Offering a Balance Between Quality and Speed: Efficient utilization of resources while maintaining image quality.**

CHAPTER 3

Proposed Methodology

3.1 System Design

The **AI Image Generator** consists of a **frontend and backend** system working together to generate images based on user prompts. The **Stable Diffusion model**, running on ComfyUI, processes textual inputs and converts them into images step by step.

3.2 System Design

The system follows a modular approach, divided into:

- **Frontend: Developed using HTML, CSS, and JavaScript, allowing users to enter prompts and select image generation settings.**
- **Backend: Implemented using Python with ComfyUI, responsible for processing the input and generating images.**
- **AI Model: Stable Diffusion, which generates images by iteratively refining latent representations.**

3.3 Workflow of the System

1. **User Input: The user enters a textual prompt in the UI.**
2. **Request Handling: The frontend sends a request to the backend.**
3. **Image Processing:**
    - **The backend initializes the Stable Diffusion pipeline.**
    - **The model iterates through multiple steps to refine the image.**

- **The processing time varies between 7 to 10 minutes.**

4. **Image Output: The generated image is sent back to the frontend for display.**

3.4 Hardware and Software Requirements

3.4.1 Hardware Requirements

Component	Specifications
Processor	Intel i5 or higher / AMD Ryzen 5 or higher
RAM	Minimum 16GB
GPU	NVIDIA RTX 3060 or higher (recommended for faster processing)
Storage	50GB free space for model and outputs

3.4.2 Software Requirements

Software	Description
Python	Backend programming language (ComfyUI support)

HTML,CSS, **JavaScript**	Frontend development
Stable diffusion model	AI based image generation
Comfy UI	Modular processing framework

3.5 Advantages of Using ComfyUI

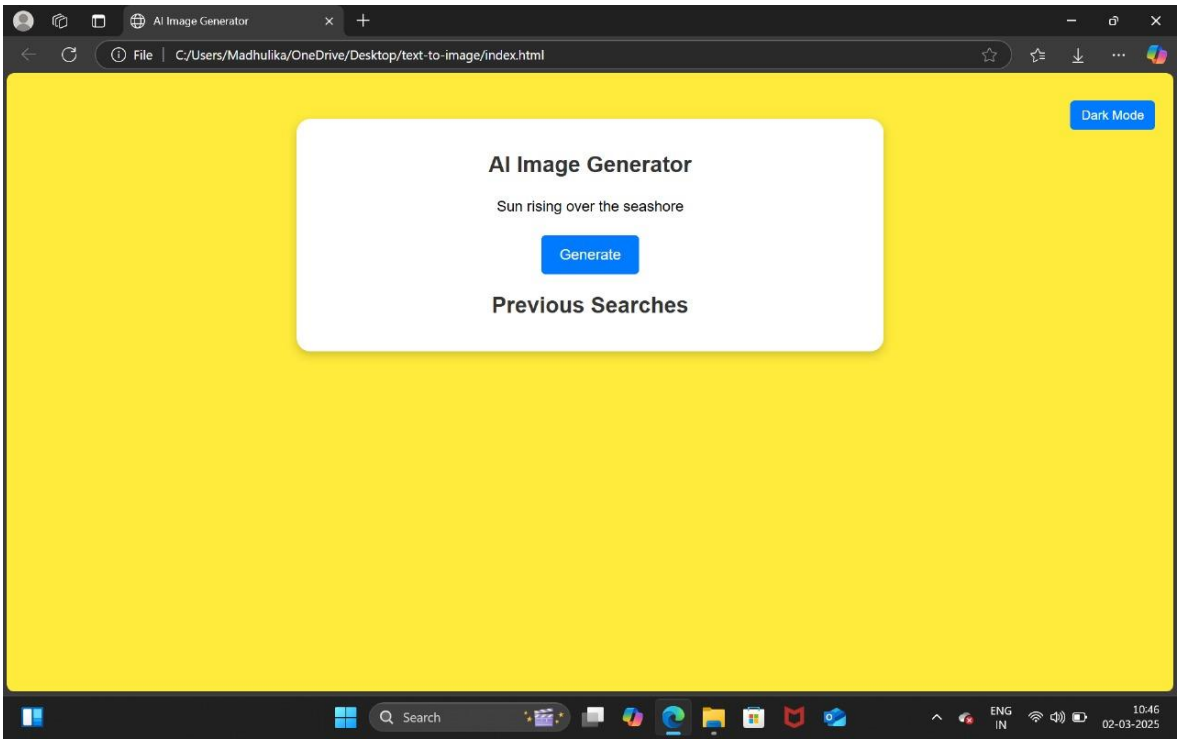
- **Step-by-Step Execution**: Provides better control over image generation.
- **Customizability**: Allows fine-tuning of AI-generated images.
- **Efficiency**: Reduces processing overhead compared to alternative interfaces.

CHAPTER 4

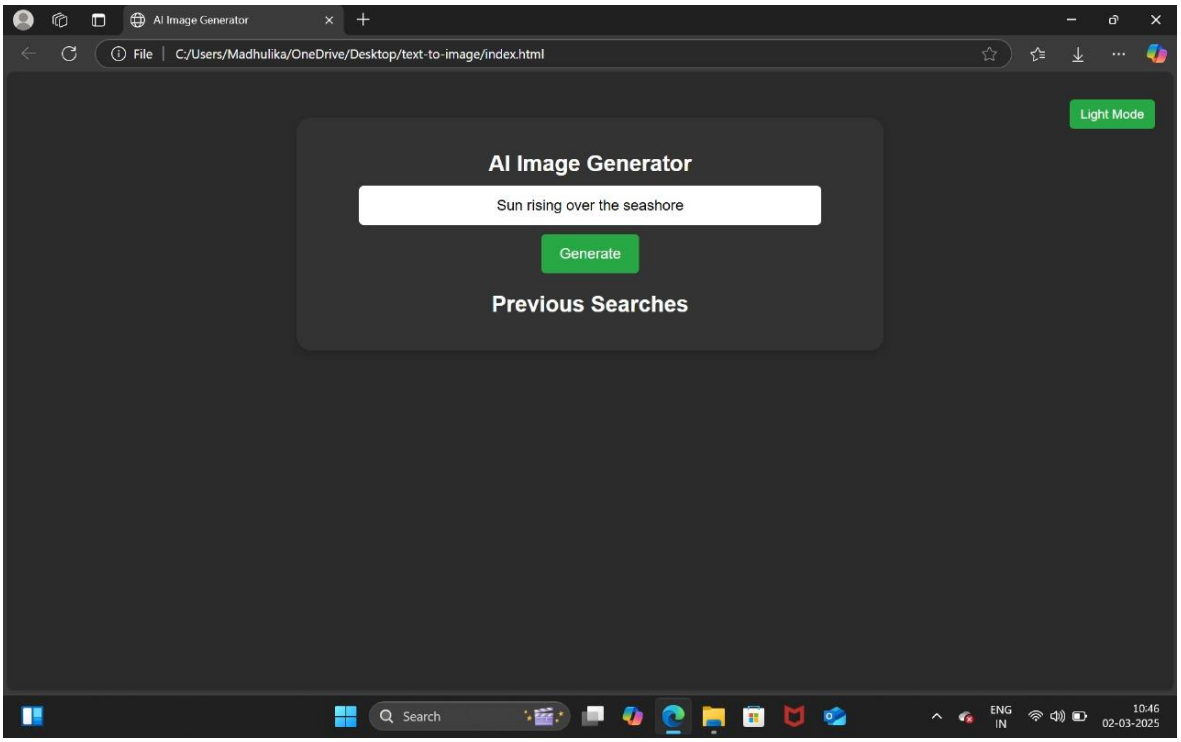
Implementation and Result

4.1 Snap Shots of Result:

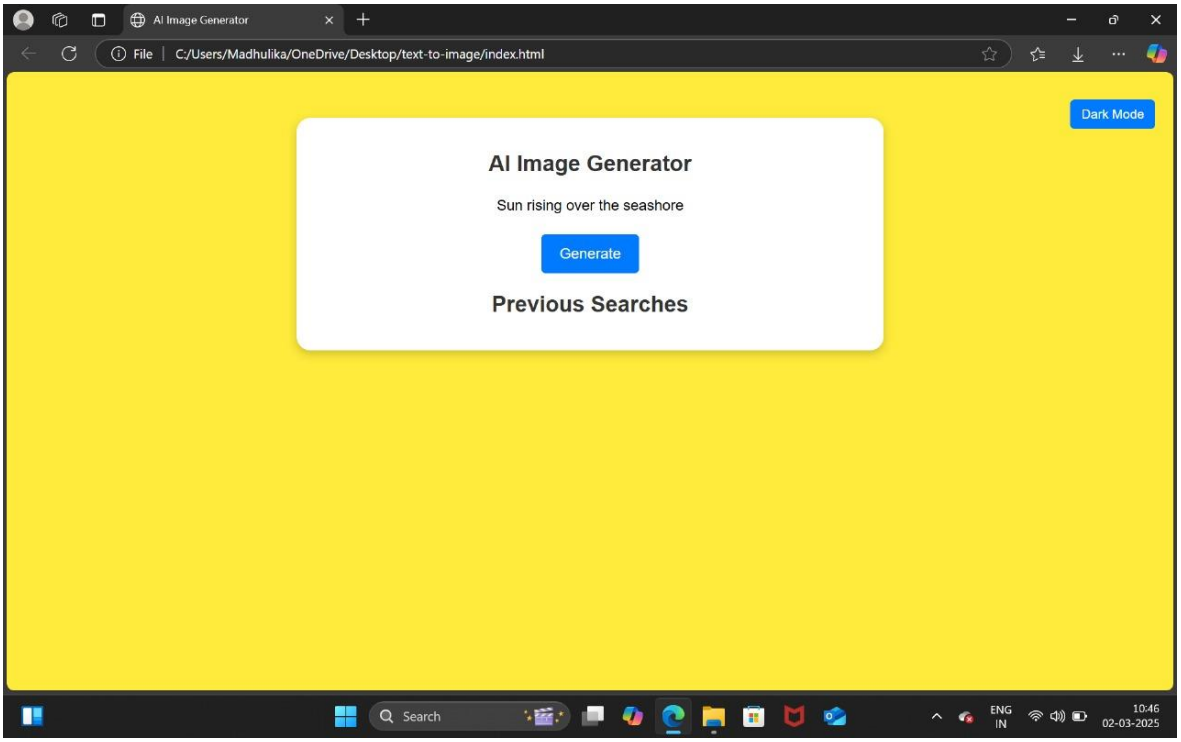
1.User Interface (Light Mode): Showcasing the basic interface in light mode with a centered heading and input prompt bar.



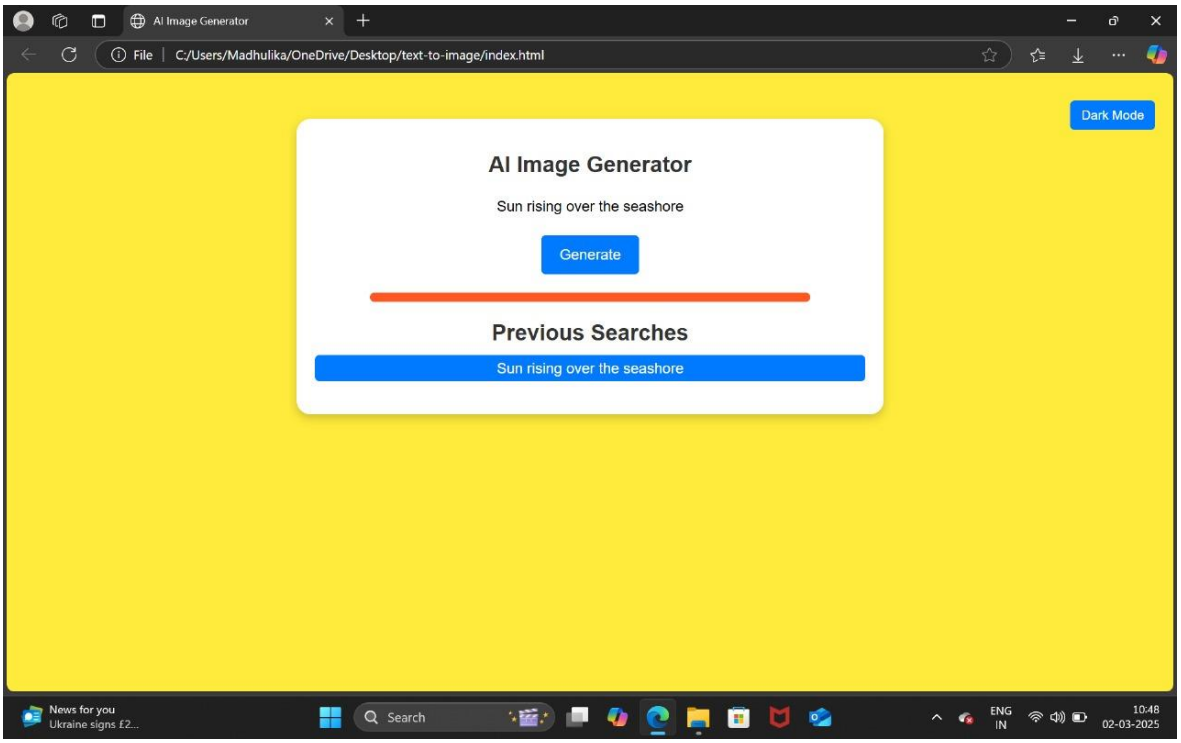
2.User Interface (Dark Mode): Displaying the dark mode variant for better accessibility.



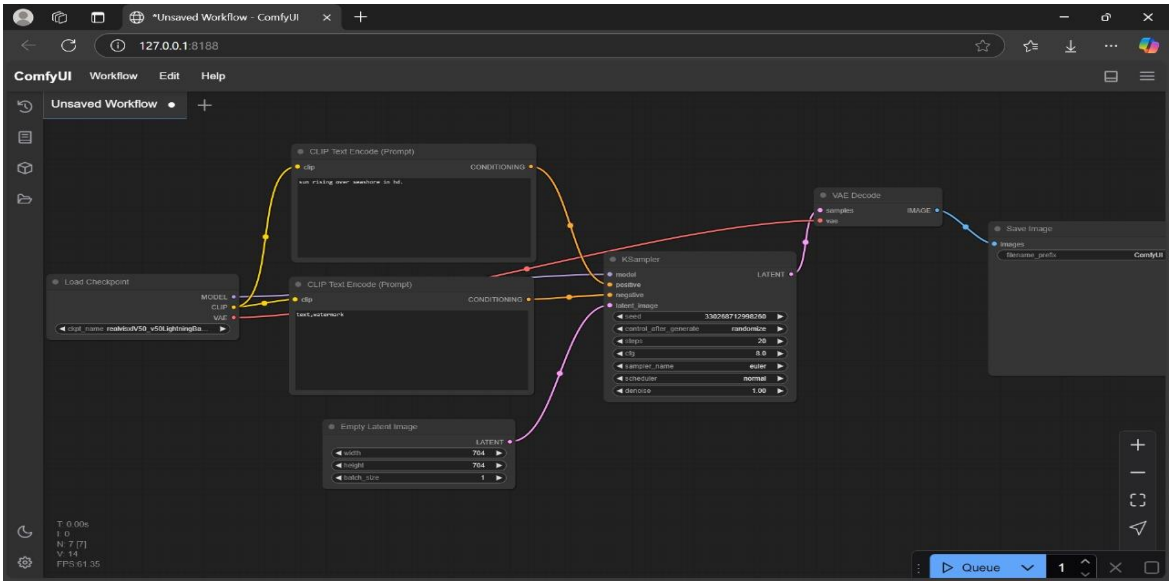
3.Giving Prompt: Demonstrating how a user enters a prompt for image generation.



4.Generating Loading Bar: Visual representation of the system processing the request.

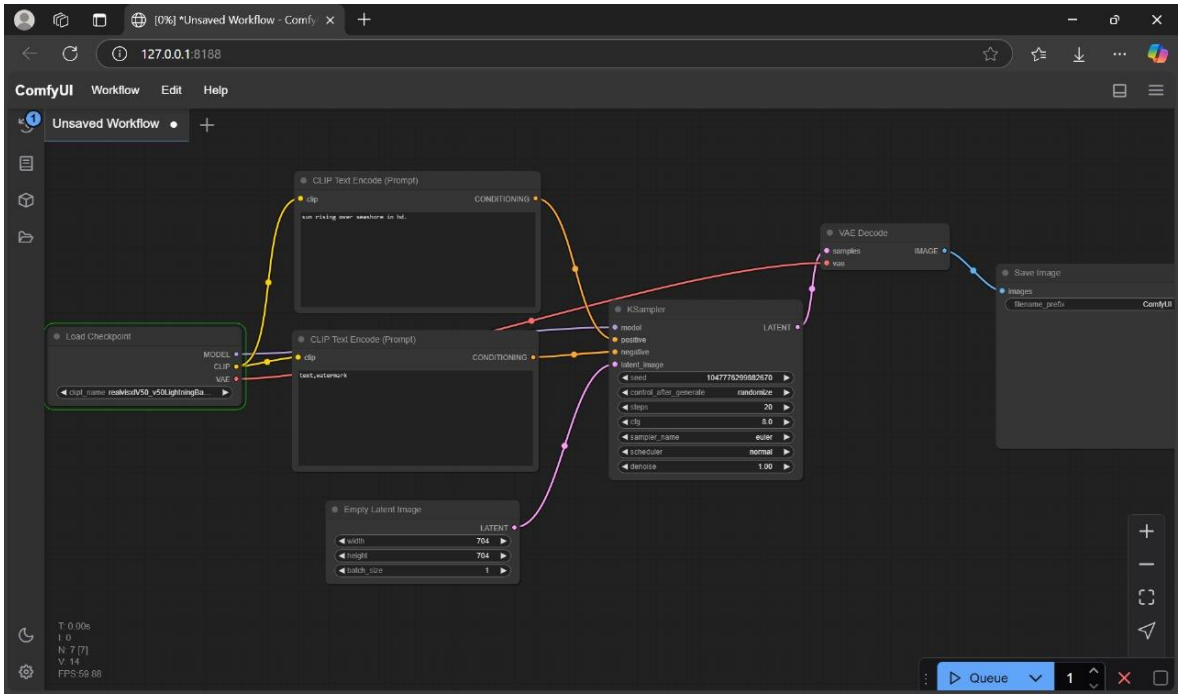


5.Structure of ComfyUI: Overview of how ComfyUI is structured for processing

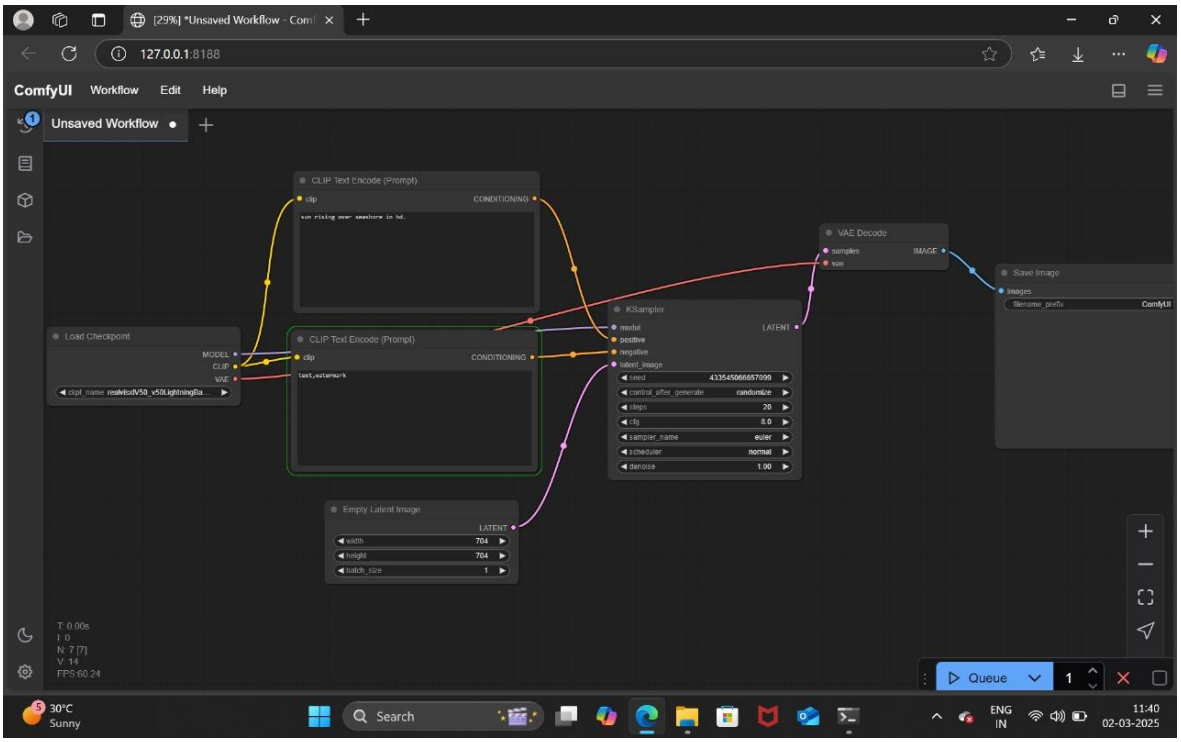


6.Step-by-Step Process in ComfyUI: Explaining how the image generation is carried out in steps.

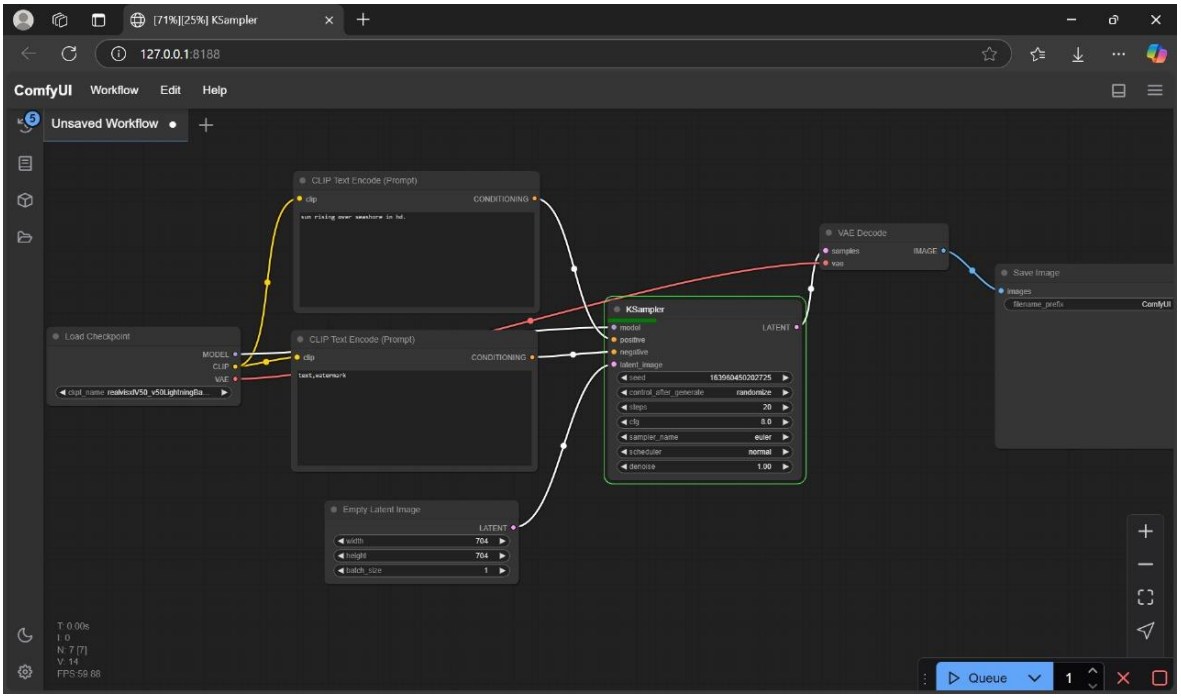
6.1 Initial step of image generation using diffusion package. Selecting the diffusion package at checkpoints and click the button Queue, it will start the process.



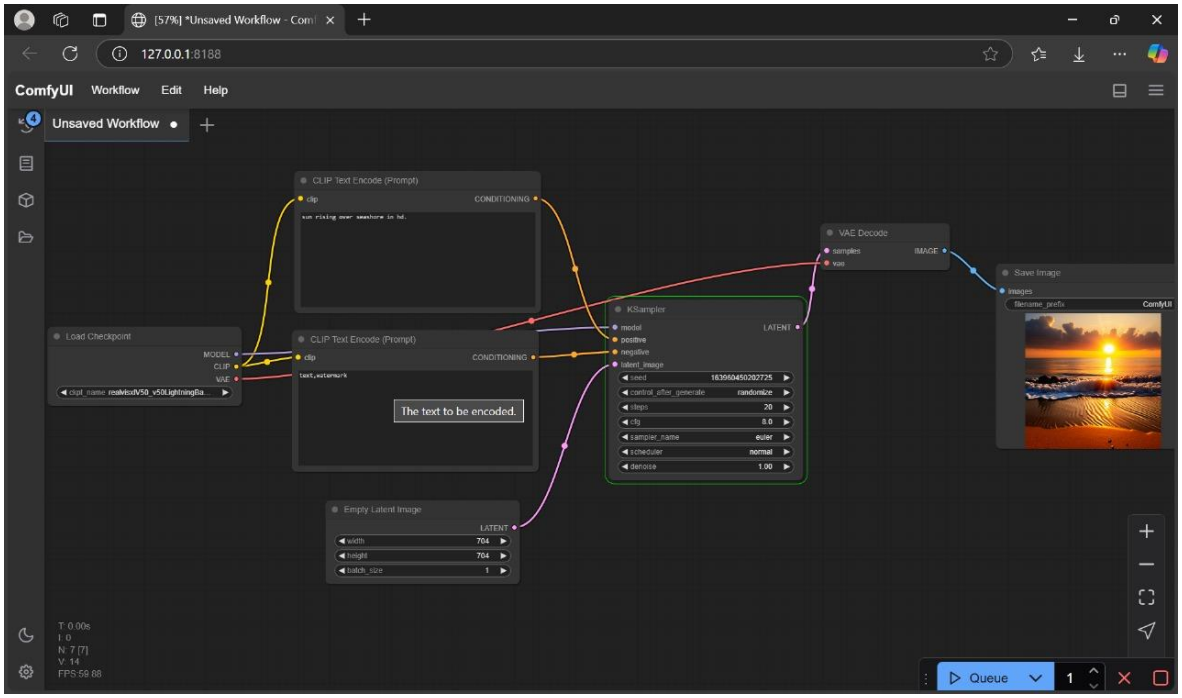
6.2 Then it collects the given prompts from the "First CLIP Text Encode(Prompt)" as per the default constraints, which are in "Second CLIP Text Encode(Prompt)" and sends to the Ksampler module.



6.3 This is the Ksampler module, which shows the loading bar that indicates the completion of the image.

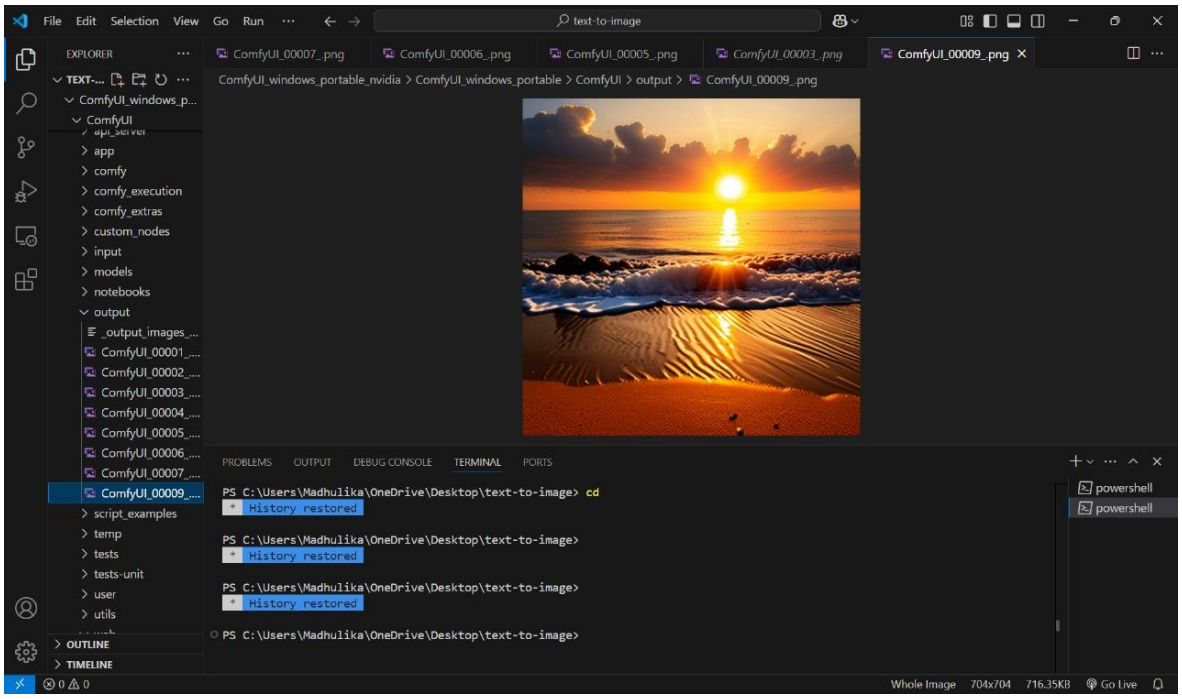


6.4 This is the VAE module it receives the signals from Ksampler and it transmits to the save image and it will generate the image.

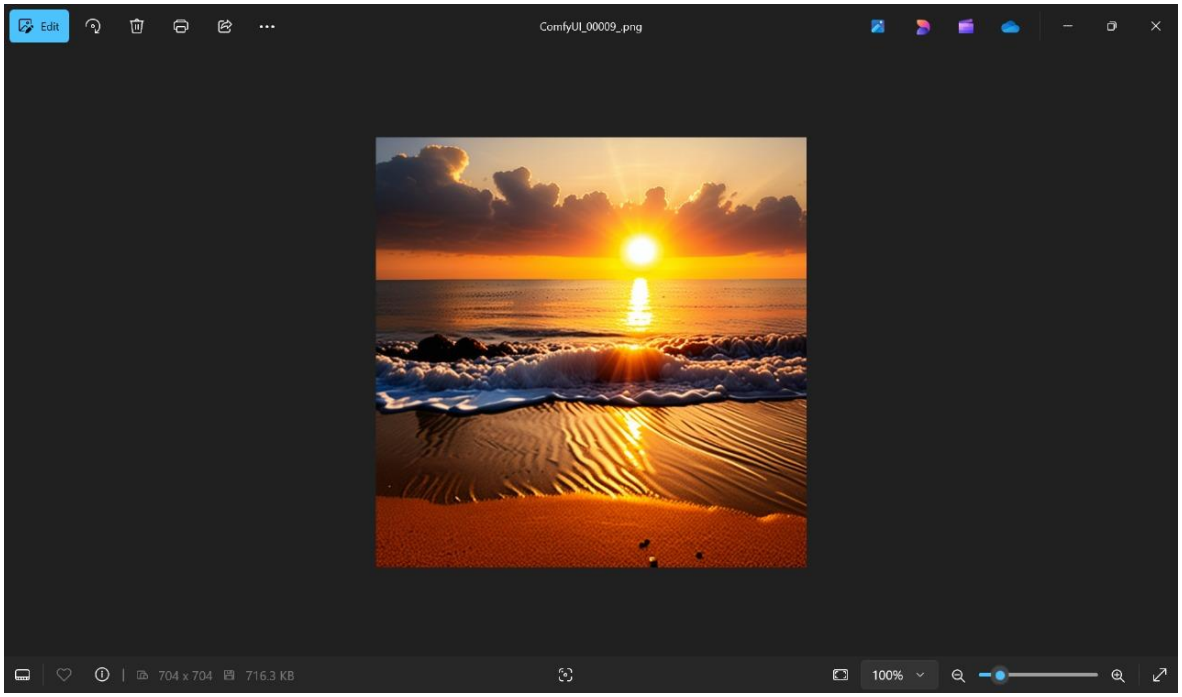


7.Stored Images: All the generated images are stored in a separate folder.

S



8.Final Image: This is the final generated image after processing.



CHAPTER 5

Discussion and Conclusion

5.1Future Work:

The project lays the foundation for future enhancements in AI-driven image generation. Some possible areas for improvement include:

5.2.1 Faster Image Generation

- Implementing optimized AI models (such as SDXL or ControlNet) for reducing generation time.
- Parallel processing techniques for more efficient handling of image requests.

5.2.2 Cloud-Based Deployment

- Hosting the AI model on cloud services such as Google Cloud, AWS, or Hugging Face Spaces to reduce local computational requirements.
- Providing API-based image generation, allowing multiple users to access the model simultaneously.

5.2.3 Enhanced User Controls

- Allowing users to adjust parameters (e.g., resolution, image style, color tone) for more customized image outputs.
- Implementing real-time previews before final image generation.

5.2.4 Expansion to Other AI Models

- Incorporating GAN-based models (e.g., StyleGAN) for additional artistic variations.

- Exploring multimodal AI to combine text and sketch-based image generation.

5.3 Conclusion

The **AI Image Generator** successfully demonstrates how **Stable Diffusion and Comfy UI** can be used to generate high-quality images based on **text prompts**. By designing an **intuitive UI** and **structured backend**, the project provides a seamless way for users to interact with AI-powered creativity.

While the current system **requires long processing times and high computational power**, future improvements in **hardware acceleration, cloud integration, and model optimization** can help make AI-based image generation **faster, more efficient, and accessible to a wider audience**.

This project serves as a **strong foundation** for exploring **AI-driven art, content creation, and automated design tools**, paving the way for **future advancements in generative AI**.

REFERENCES

1. Rombach, Robin, et al. *High-Resolution Image Synthesis with Latent Diffusion Models.* CVPR 2022.
2. Ho, Jonathan, et al. *Denoising Diffusion Probabilistic Models.* NeurIPS 2020.
3. Official documentation of Stable Diffusion and ComfyUI framework.
4. AI image generation research papers and relevant online resources.