**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

**Sehajpreet Kaur**

Under the Guidance of

**Jay Rathod and Adharsh P**

**ACKNOWLEDGEMENT**

Firstly, I would like to express my heartfelt gratitude to my supervisors, Adarsh P and Jay Rathod Sir, for their invaluable mentorship and guidance throughout this journey. Their advice, encouragement, and constructive feedback have been instrumental in shaping innovative ideas and driving the successful completion of this project.

The confidence they placed in me served as a constant source of motivation. It has been an honor to work under their guidance. Their unwavering support extended beyond this project, enriching my learning experience in many aspects of the program.

Their insights and lessons have not only contributed to my project and academic growth but have also played a significant role in shaping me into a more responsible and professional individual.

#### **ABSTRACT**

The advancement of AI in image generation has revolutionized creative design, content creation, and artistic exploration. This project leverages Stable Diffusion, a state-of-the-art deep learning model, with ComfyUI, a modular and user-friendly interface, to generate high-quality images from textual descriptions.

A key challenge in AI image generation is achieving realistic, detailed, and customizable outputs without extensive computational expertise. Traditional methods often lack fine control and require significant processing power. This project aims to streamline the process, making it more accessible and efficient.

The objectives include implementing Stable Diffusion for high-fidelity image generation, enhancing control through prompt engineering, and optimizing model parameters for better results. By integrating ComfyUI, users can easily modify prompts, adjust settings, and experiment with diffusion parameters to refine their outputs.

The methodology involves setting up Stable Diffusion, designing effective prompts, fine-tuning parameters such as diffusion steps and guidance scale, and evaluating images based on realism, coherence, and user satisfaction. Systematic experimentation demonstrates that strategic prompt engineering and parameter adjustments significantly improve output quality and diversity.

Results indicate that Stable Diffusion with ComfyUI offers a powerful, flexible, and efficient approach to AI-driven image creation. The project highlights the importance of user interaction, customization, and optimization in generating high-quality visuals. These findings lay the groundwork for further exploration in digital art, advertising, entertainment, and automated design workflows.

**TABLE OF CONTENT**

**Abstract I**

**Chapter 1.**  **Introduction 1**

1.1 Problem Statement 1

1.2 Motivation 1

1.3 Objectives 2

1.4. Scope of the Project 1

**Chapter 2.**  **Literature Survey 2**

**Chapter 3.**  **Proposed Methodology 4**

**Chapter 4.**  **Implementation and Results 6**

**Chapter 5. Discussion and Conclusion 13**

**References** 15

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Caption** | **Page No.** |
|  | System Architecture |  |
|  | Initiation of the project on command prompt |  |
|  | Cyberpunk Cityscape at Night |  |
|  | Mystical Library of Lost Knowledge |  |
|  | Underwater Kingdom of Bioluminescent Creatures |  |
|  | AI-Powered Steampunk Robot Laboratory |  |
|  | Fantasy Warrior in a Battle Against a Shadow Dragon |  |
|  | Post-Apocalyptic Desert Nomad Camp |  |
|  | Dreamlike Floating Islands in the Sky |  |
|  | Gothic Vampire Castle at Full Moon |  |
|  | Futuristic Samurai in a Neon-Drenched Tokyo |  |
|  | Whimsical Candyland with Talking Creatures |  |
|  | Prompt Execution Time Details |  |

**CHAPTER 1**

**Introduction**

* 1. **Problem Statement:**

The rapid advancement of artificial intelligence in image generation has revolutionized digital content creation. However, traditional image generation methods often struggle with producing high-resolution, realistic, and customizable images. Many existing AI models require extensive computational resources, complex coding knowledge, and lack user-friendly interfaces, making them inaccessible to non-experts. Additionally, achieving precise control over image attributes, such as style, composition, and fine details, remains a significant challenge. This project addresses these issues by utilizing Stable Diffusion for high-quality AI-generated images and integrating ComfyUI, a modular and visual workflow interface, to enhance user accessibility and customization.

* 1. **Motivation:**

The project was chosen due to the growing demand for AI-driven creative tools in various industries, including digital art, game design, advertising, and content creation. AI-generated images have vast applications in automating design tasks, reducing production costs, and enabling new artistic possibilities. This approach has the potential to bridge the gap between AI technology and creative professionals, empowering users to generate high-quality visuals without deep technical expertise.

* 1. **Objective:**

The main objectives of this project are:

* To implement Stable Diffusion for generating high-quality AI-driven images.
* To integrate ComfyUI for a user-friendly and modular image generation workflow.
* To explore and optimize prompt engineering techniques for enhanced image realism and control.
* To analyse and evaluate generated images based on realism, diversity, and coherence.
  1. **Scope of the Project:**

The project focuses on the following key areas:

* AI Image Generation: Using Stable Diffusion to create images from text prompts.
* User Interface: Utilizing ComfyUI to simplify the image generation process.
* Customization & Control: Exploring techniques to fine-tune image attributes through prompt engineering and parameter adjustments.

**CHAPTER 2**

**Literature Survey**

* 1. **Review relevant literature or previous work in this domain.**

**Paper 1: “IMAGE GENERATION WITH STABLE DIFFUSION AI” By Sasirajan, Guhan, Mary Reni, Maheswari, Roselin Mary**

* This AI-powered method improves suspect identification by combining Stable Diffusion with a modular system architecture, which lessens the need for conventional sketching and boosts the effectiveness of criminal investigations. This technique is a useful tool for law enforcement organizations since it can produce realistic and detailed pictures based on text descriptions.

**Paper 2: “A Comparative Analysis of AI Image Generation Models: Stable Diffusion, Dall-E and Dream by WOMBO” By Mudith Shetty, Haafiz Sheikh, Pari Sharma, Kshitij Shrivastava, Jesleena Gonsalves**

* AI-powered picture creation has advanced dramatically, with models such as WOMBO's Dream, DALL-E, and Stable Diffusion showcasing special characteristics. These models solve issues with originality, realism, and computing efficiency by using deep learning algorithms to convert textual descriptions into high-quality visuals.

**Paper 3: “Stable Diffusion Model: The Past, Present and the Future in India” By Atharv Singh**

* The paper highlights India's quick progress in AI while also drawing attention to serious issues with data privacy, moral obligation, and governmental supervision. Comprehensive regulatory frameworks are urgently needed to ensure fair, transparent, and responsible AI systems as AI is incorporated into more and more areas. Furthermore, promoting trust and ethical adoption of AI requires increasing public knowledge of its effects, advantages, and risks. Future studies should concentrate on improving AI safety, reducing prejudice, and making sure ethical standards are followed. For AI to be deployed in India in a sustainable and responsible manner, a balanced strategy that incorporates innovation and regulation is necessary.
  1. **Mention any existing models, techniques, or methodologies related to the problem.**

Numerous AI-powered methods improve facial recognition, suspect identification, and text-to-image generation.

**Models for text to images**

* Stable Diffusion: Finely controls the generation of detailed visuals from text.
* DALL-E: Constructs imaginative pictures from written descriptions.
* DeepFace Drawing: Realistic facial images can be created from sketches.

**Recognition and Reconstruction of the Face**

* GANs: Improve face reconstruction.
* StyleGAN: Produces lifelike human faces.
* Eigenfaces and Fisherfaces: PCA and LDA-based traditional facial recognition.

**Systems for Suspect Identification**

* Sketch-to-Image for Forensic AI: Creates lifelike visuals from sketches.
* DeepFace: High-dimensional facial identity mapping.
* Composite Face Recognition: Improves accuracy by combining AI models.
  1. **Highlight the gaps or limitations in existing solutions and how your project will address them.**

|  |  |
| --- | --- |
| **Existing Limitations** | **Proposed Solutions** |
| **High computational requirements for GAN-based models.** | **Utilize Stable Diffusion, which is more efficient and requires lower computational resources.** |
| **Lack of user-friendly control over image attributes in many models.** | **Implement ComfyUI, allowing intuitive node-based customization of image generation.** |
| **Difficulties in achieving precise customization using text prompts alone.** | **Optimize prompt engineering and parameter tuning to enhance image coherence and control.** |
| **Existing solutions require deep technical expertise.** | **ComfyUI provides a visual, drag-and-drop workflow, making AI image generation accessible to non-experts.** |
| **Many models lack flexibility for iterative experimentation.** | **Enable real-time modifications using ComfyUI, allowing users to refine image generation efficiently.** |

**CHAPTER 3**

**Proposed Methodology**

* 1. **System Design**

**Figure 1: System Architecture**

* Load Checkpoint: Loads a pre-trained model checkpoint for image generation.
* Clip Text Encode Positive: Encodes positive text prompts for image generation.
* Clip Text Encode Negative: Encodes negative text prompts to guide the image generation process.
* Empty Latent Image: Creates an empty latent space for generating new images.
* K Sampler: Samples from latent space using the K-means algorithm to refine generated images.
* VAE Decode: Decodes latent images back into their visual representation.
* Save Image: Saves the generated image to the system’s storage.
  1. **Requirement Specification**
     1. **Hardware Requirements**
* **CPU**: Intel Core i5 (8th Gen) or AMD Ryzen 5
* **RAM**: 8GB
* **GPU**: NVIDIA GPU with at least **4GB VRAM** (e.g., GTX 1650)
* **Storage**: 20GB free SSD space
* **Operating System**: Windows 11 (64-bit)
  + 1. **Software Requirements**
* **ComfyUI**: An intuitive, modular interface that makes it simple to modify inputs, settings, and output production for AI models such as Stable Diffusion.
* **Hugging Face Stable Diffusion Model:** Using the pre-trained weights of Stable Diffusion, this robust text-to-image AI model offered on Hugging Face allows for the creation of high-quality images from text descriptions.

**CHAPTER 4**

**Implementation and Result**

**Implementation Process**

* **Setting Up the Environment**

*Install Comfy UI*: Download Comfy UI from the official GitHub repository or use pre-built installer. Set up any dependencies as required by the platform.

* **Prepare for Image Generation**

*Download the Stable Diffusion Model Weights:* Obtain the model weights for Stable Diffusion from sources like Hugging Face or the official repository. Make sure the weights are compatible with your version of Stable Diffusion.

*Prepare the Comfy UI*: Configure Comfy UI to interact with the model by specifying paths to the model weights, configuration files, and any other necessary assets.

* **Input Parameters**

Enter a text prompt in the Comfy UI interface to describe the image you want to generate. This could be anything from "a futuristic city" to "a medieval landscape."

* **Start Image Generation**

Hit the "Queue" button in Comfy UI to start the process. Monitor the progress through the interface. Depending on your hardware, it may take a few seconds to several minutes. After completion, you can view the generated image, and it might provide options to refine or re-generate the image.

* **Post-processing and Fine-tuning**

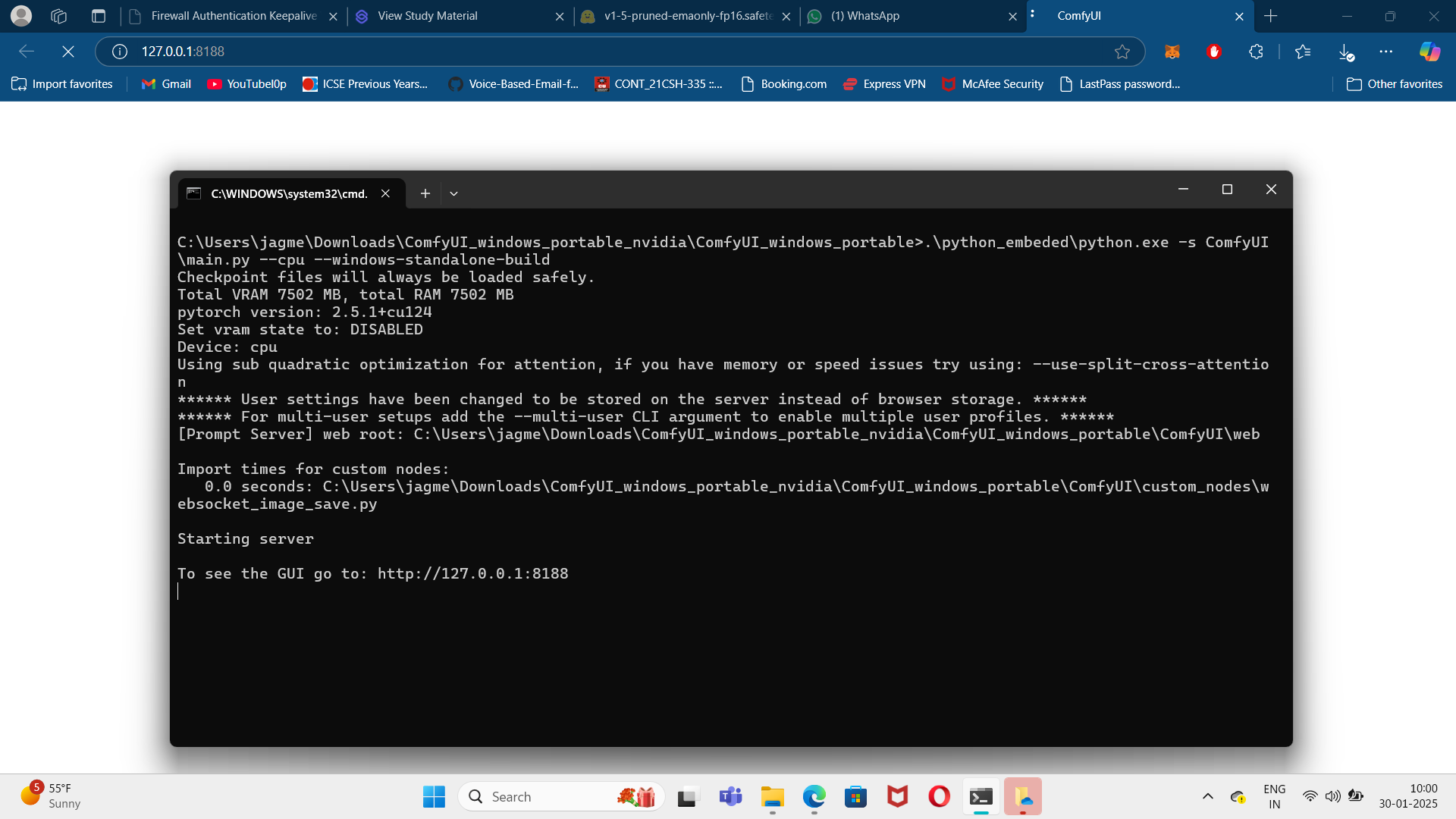
*Edit/Refine*: Use additional Comfy UI settings to fine-tune images, adjust style or details, or even apply a specific artistic filter.

*Image Enhancements*: You might also apply super-resolution models or other enhancement tools available in the platform to improve image quality.

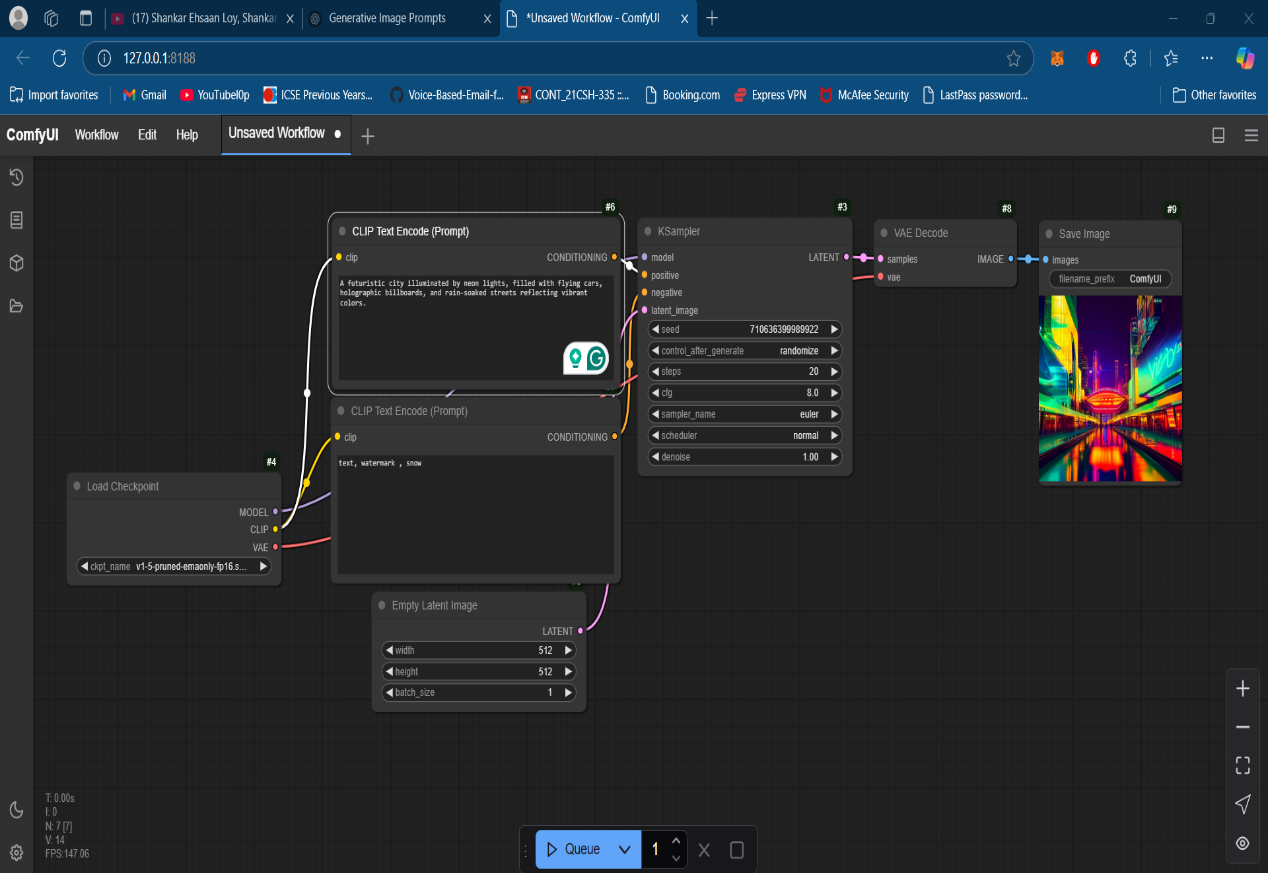
* **Save/Export:**

After generating the image, you can save it to your local directory or share it based on the platform's integration options. You can also experiment with batch generation for creating multiple images at once using different prompts or settings.

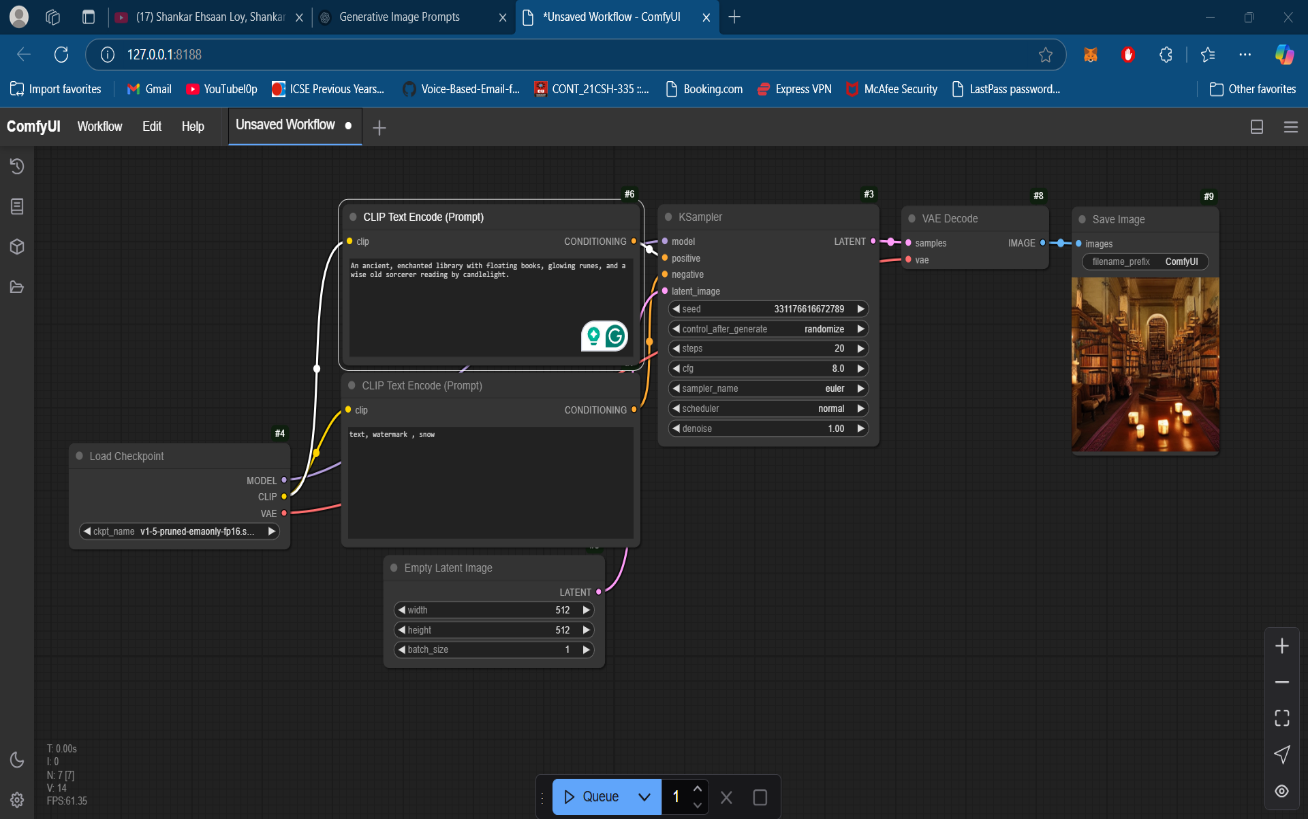
* 1. **Snap Shots of Result**

****

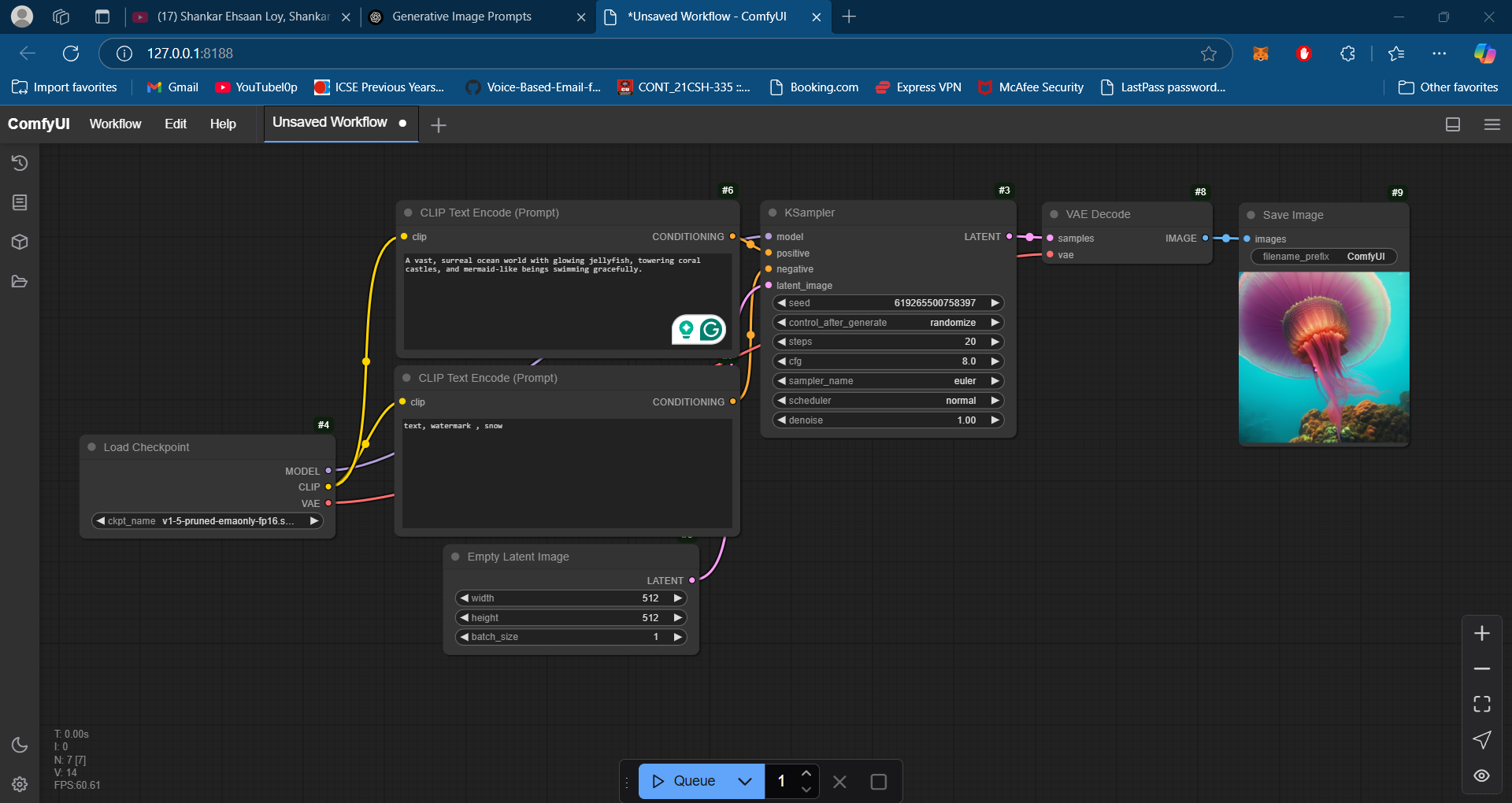
**Figure 2: Initiation of the project on command prompt**

****

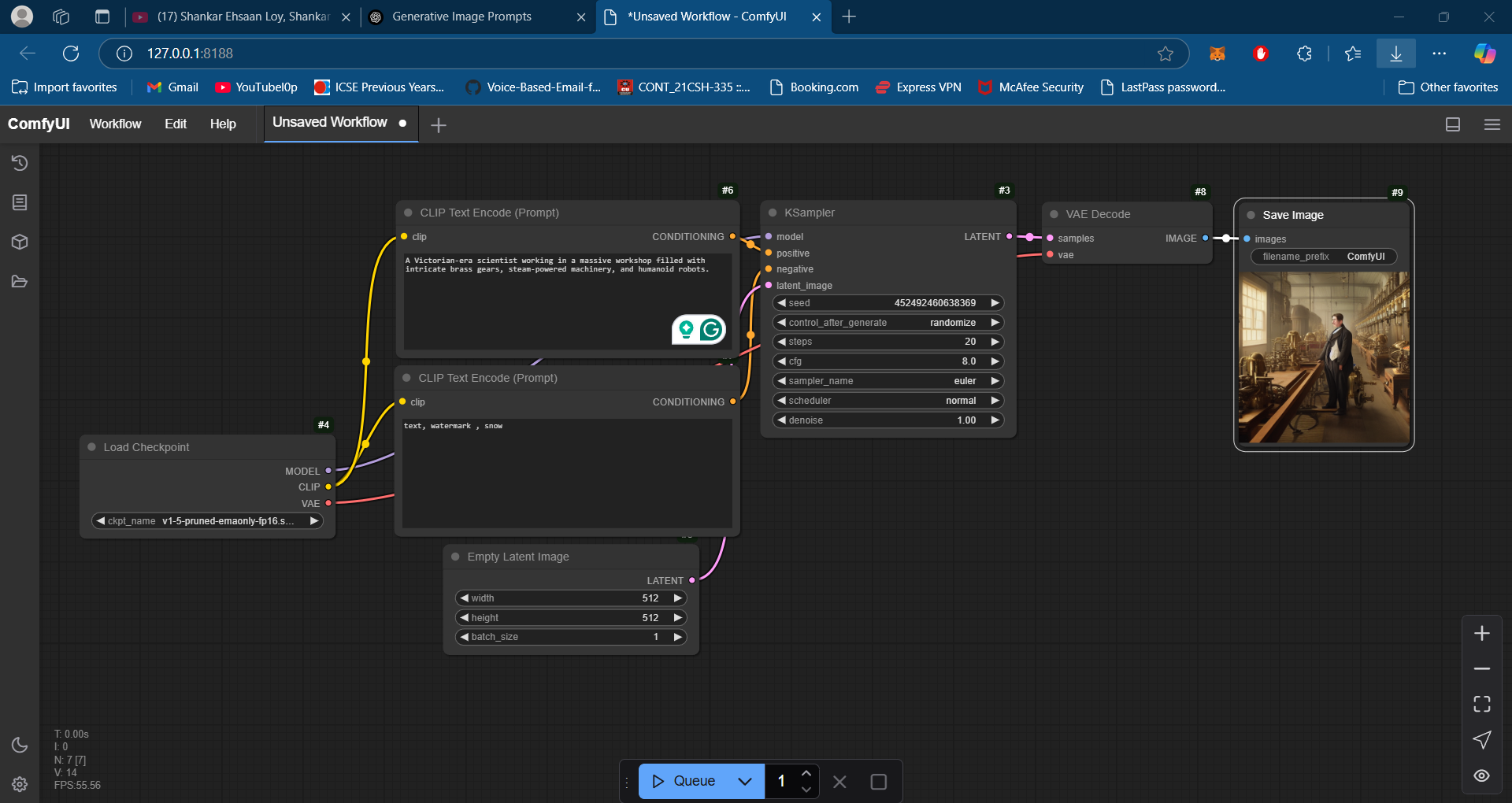
**Figure 3: Cyberpunk Cityscape at Night**

****

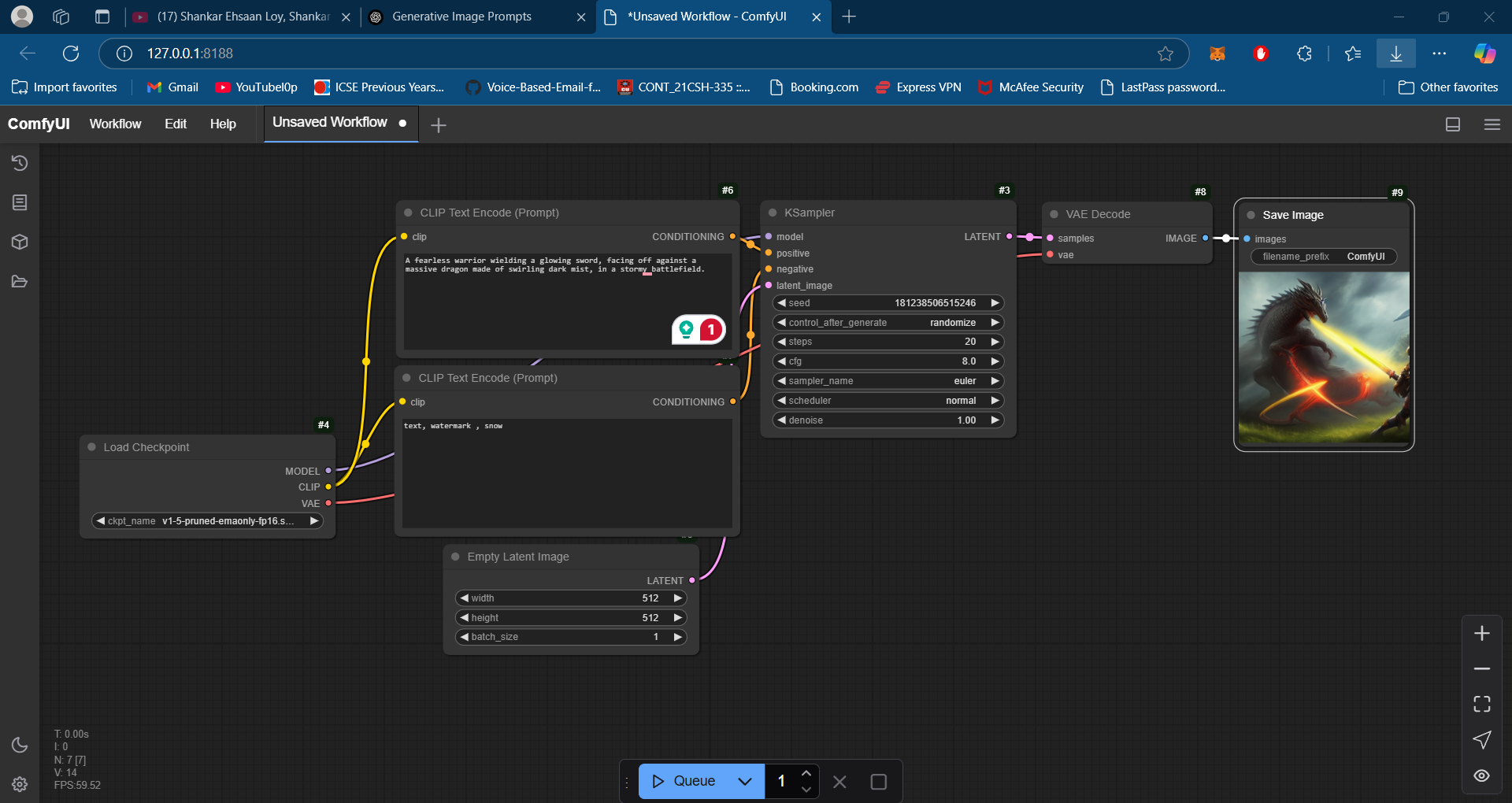
**Figure 4: Mystical Library of Lost Knowledge**

****

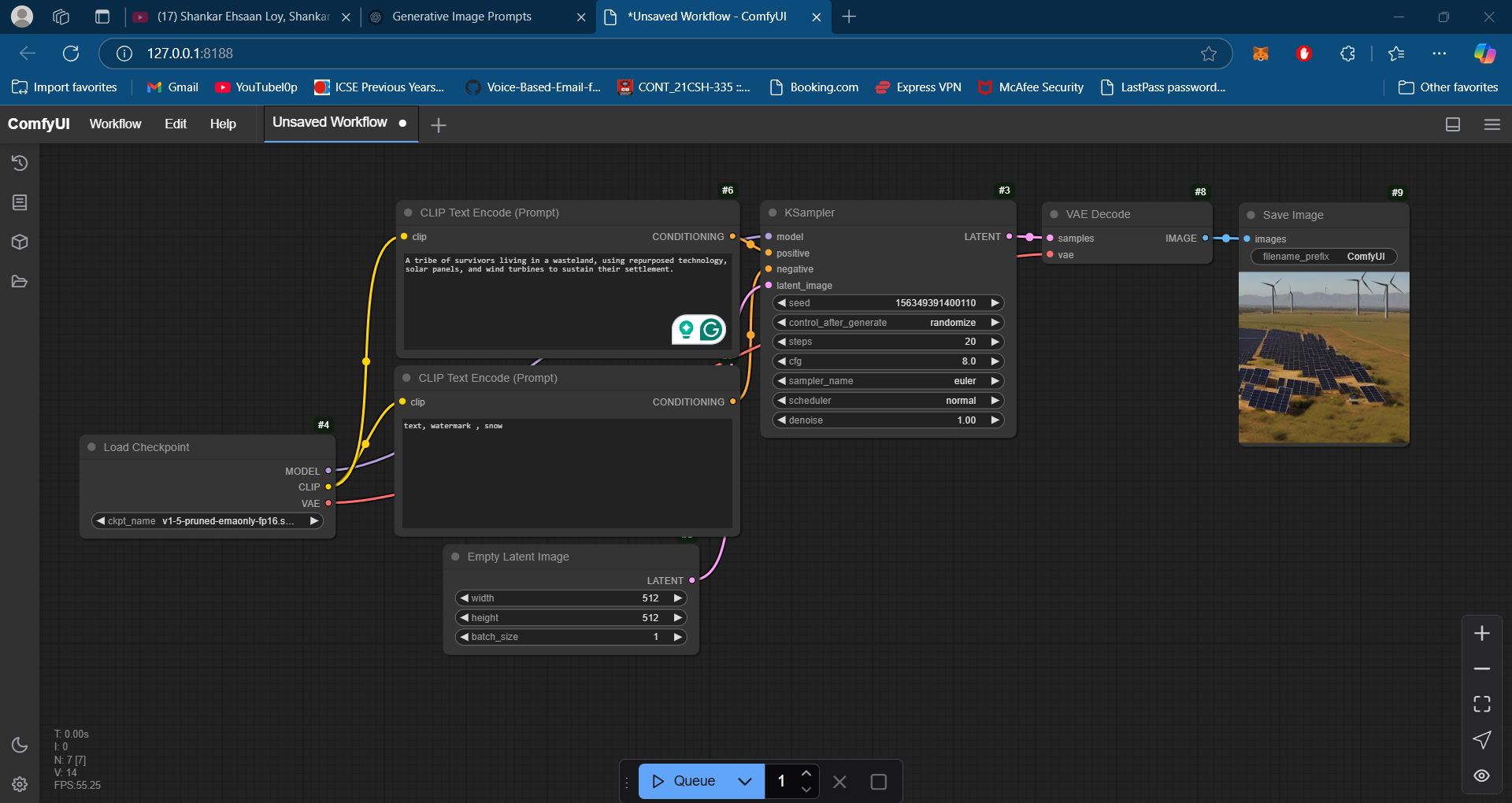
**Figure 5: Underwater Kingdom of Bioluminescent Creatures**

****

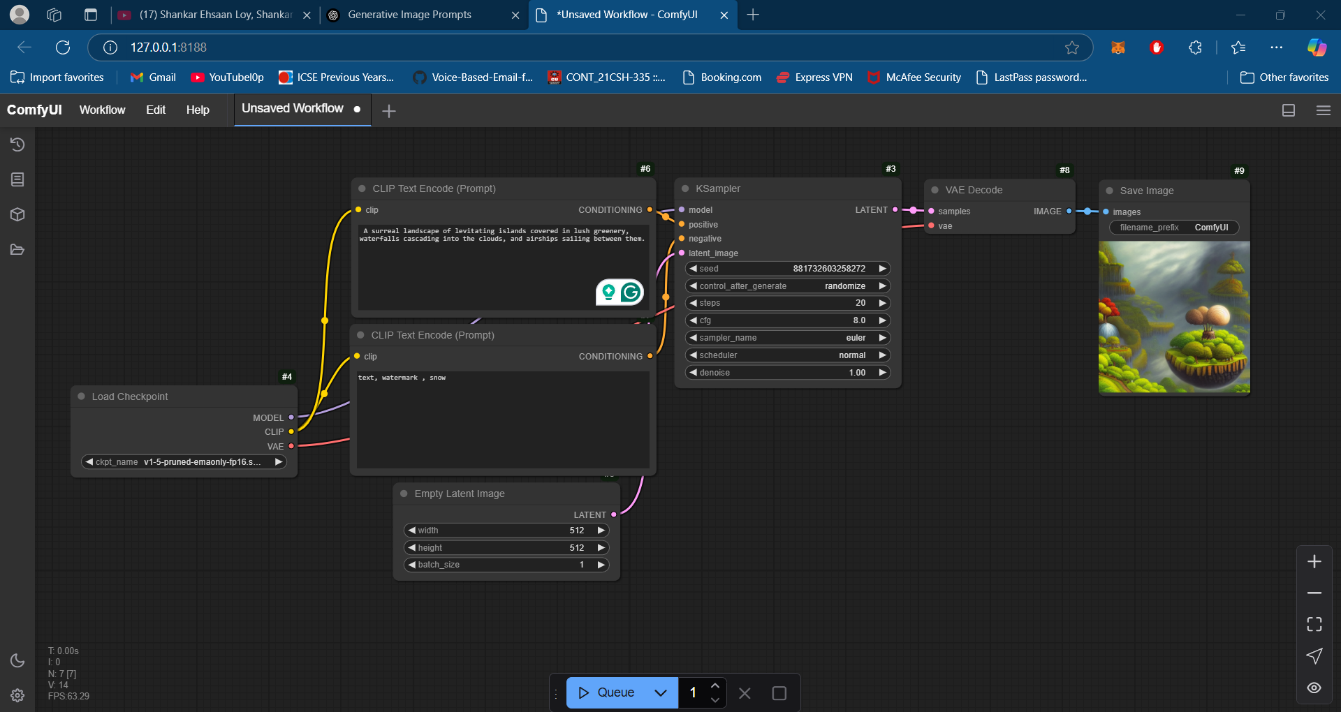
**Figure 6: AI-Powered Steampunk Robot Laboratory**

****

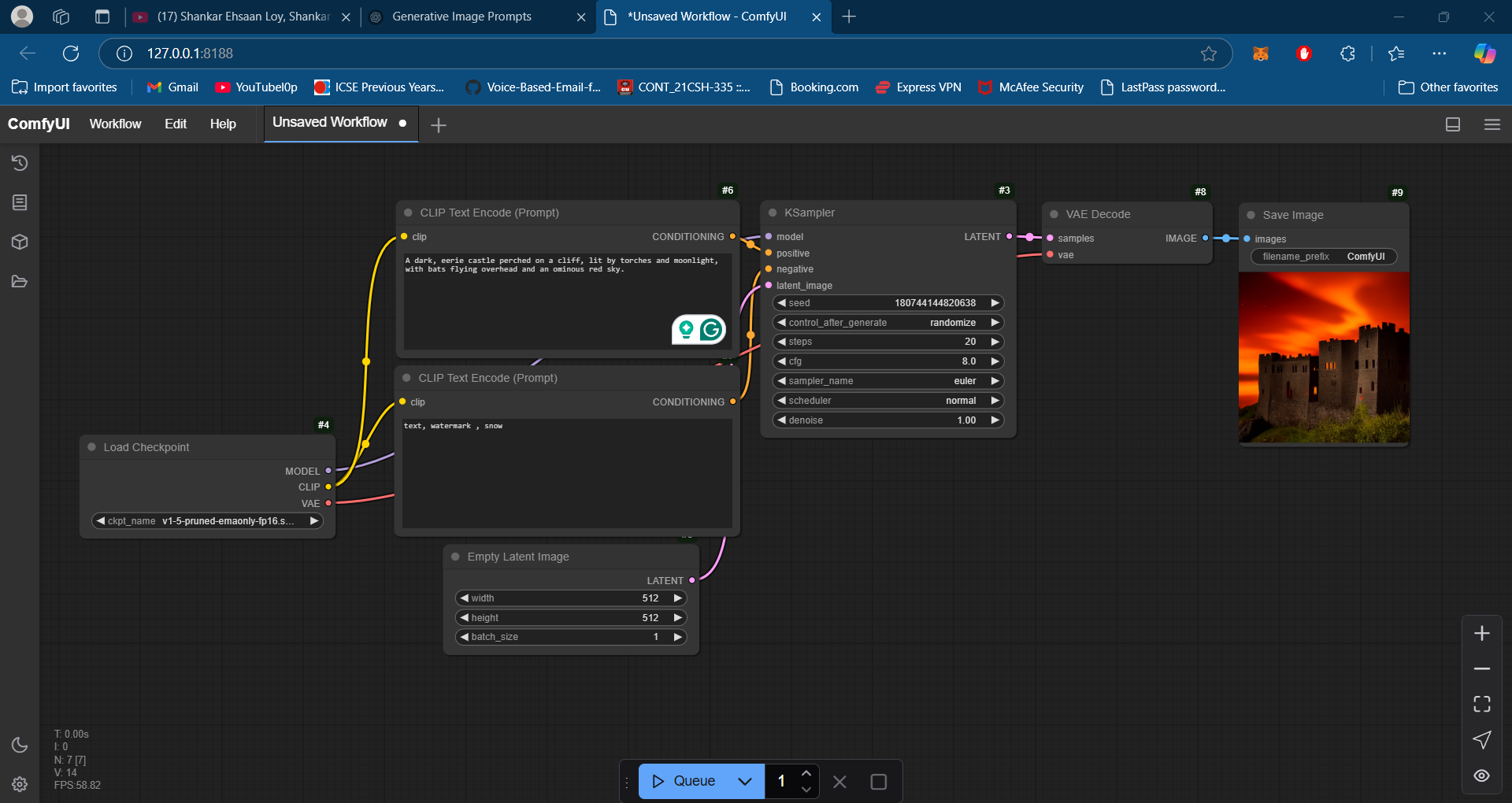
**Figure 7: Fantasy Warrior in a Battle Against a Shadow Dragon**

****

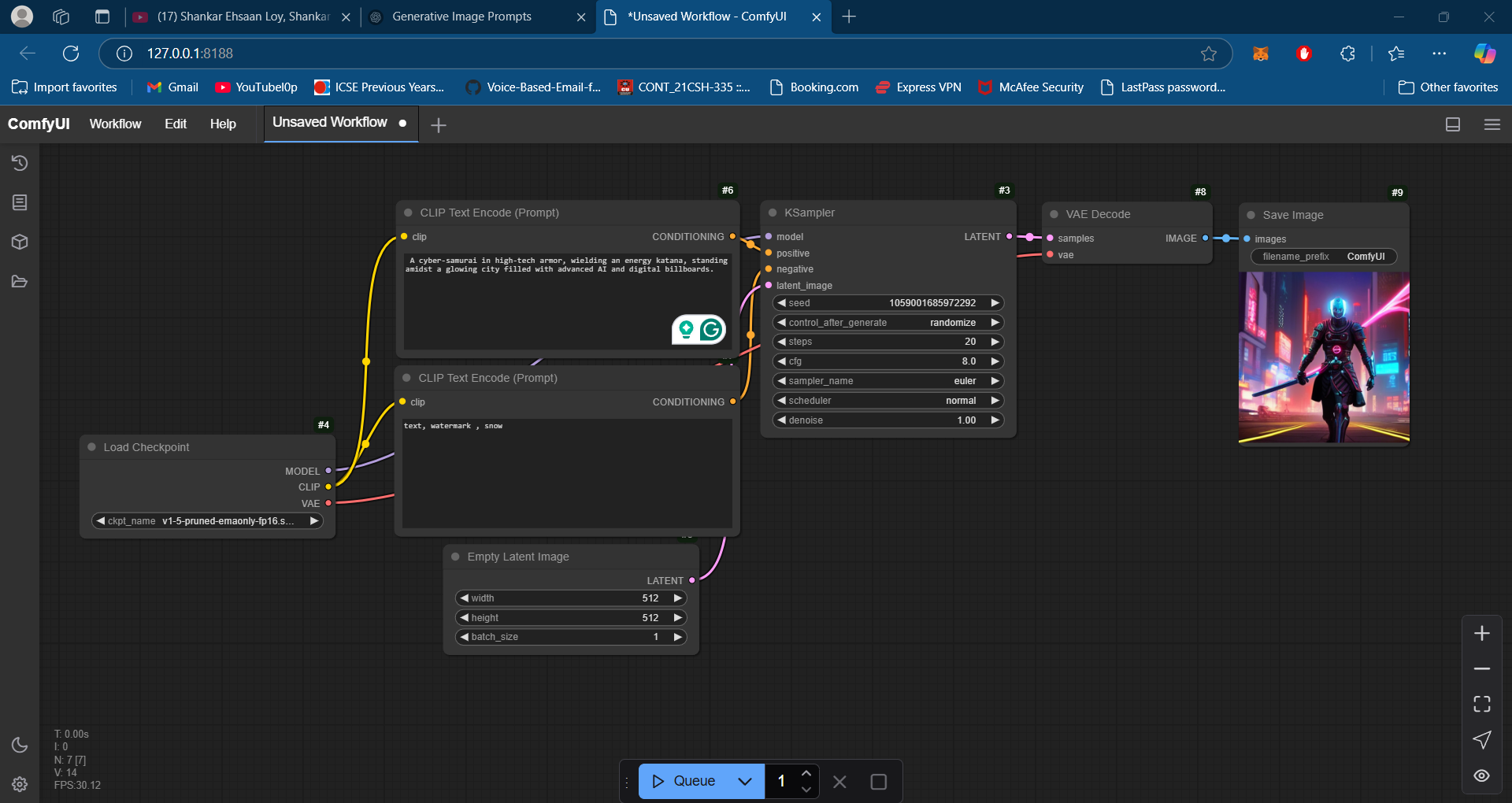
**Figure 8: Post-Apocalyptic Desert Nomad Camp**

****

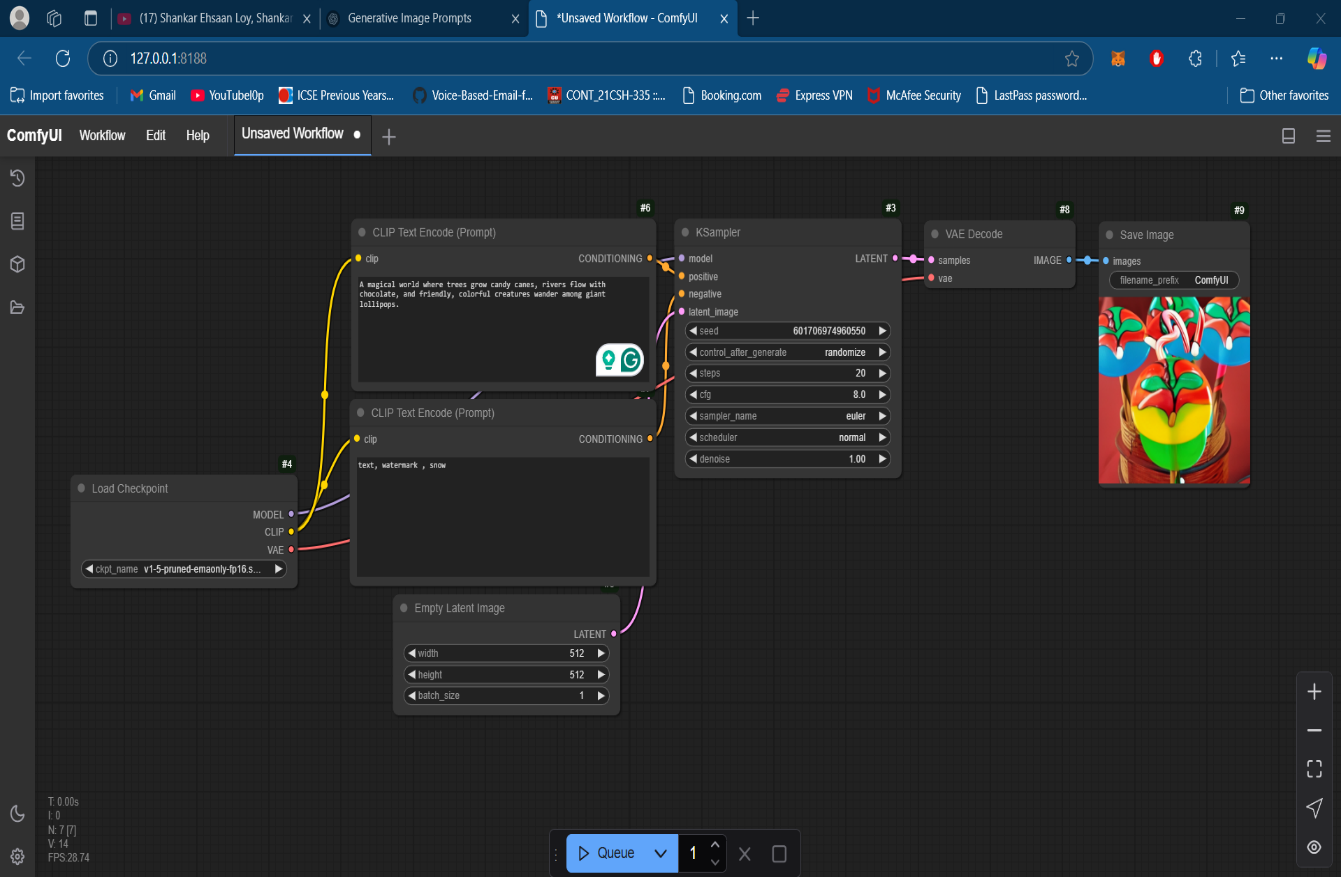
**Figure 9: Dreamlike Floating Islands in the Sky**

****

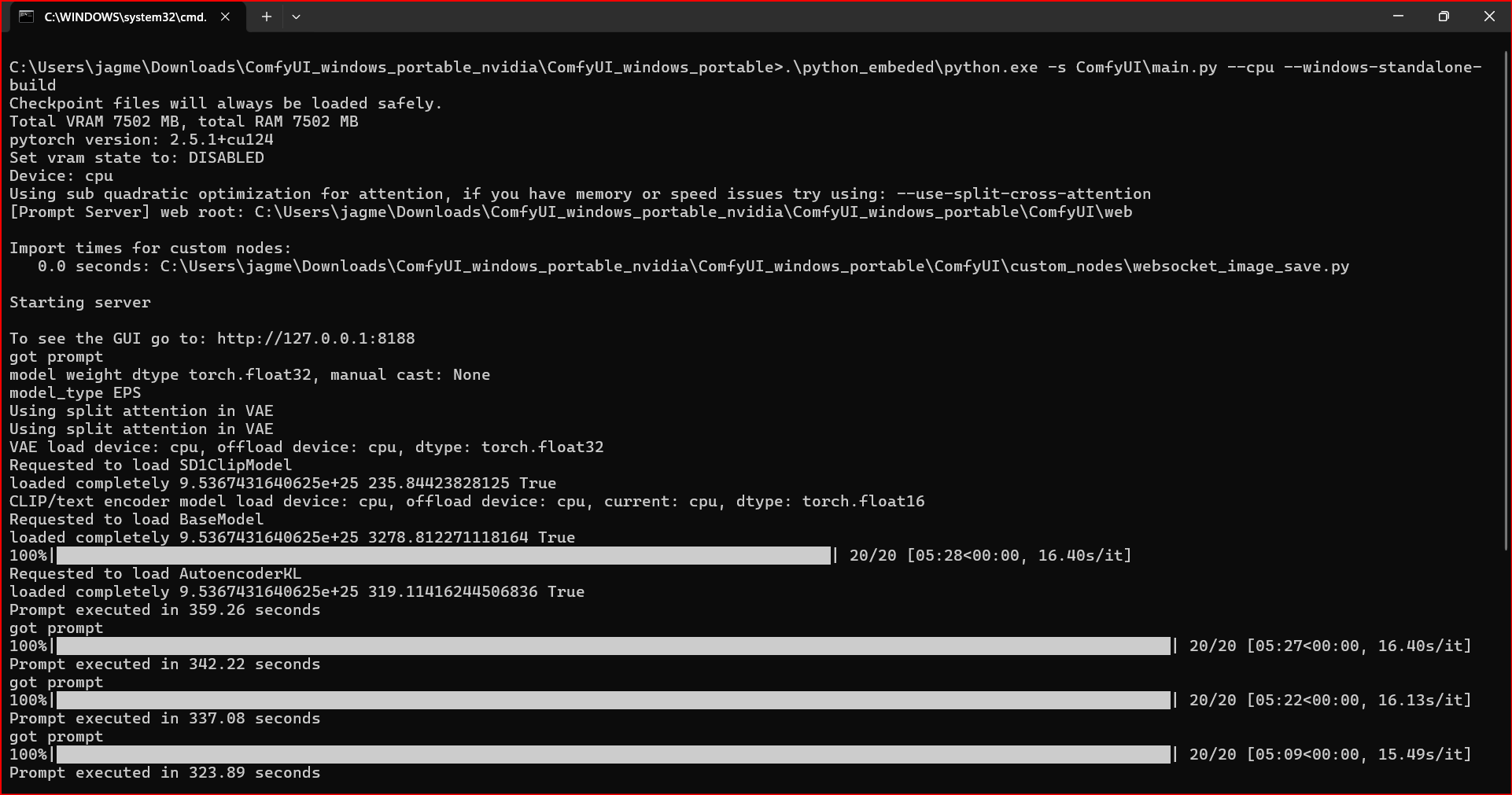
**Figure 10: Gothic Vampire Castle at Full Moon**

****

**Figure 11: Futuristic Samurai in a Neon-Drenched Tokyo**

****

**Figure 12: Whimsical Candyland with Talking Creatures**

****

**Figure 13: Prompt Execution Time Details**

**CHAPTER 5**

**Discussion and Conclusion**

* 1. **Future Work:**

To enhance the image generation model using Stable Diffusion and Comfy UI, here are some suggestions for future work:

1. **Model Optimization**: Improve the efficiency of the Stable Diffusion model to reduce processing time without sacrificing quality, possibly by exploring model pruning or quantization.
2. **Fine-tuning the Model**: Allow users to fine-tune Stable Diffusion on their specific dataset for more personalized and context-specific image generation. This could involve training the model with custom images or styles.
3. **User Interface Enhancements**: Improve the Comfy UI to support more advanced options, such as live previews, real-time adjustments during image generation, and integration with more advanced tools like inpainting or image-to-image generation.
4. **Integration with Other Models**: Expand the system by integrating it with other AI models for tasks like upscaling, object recognition, or colorization, to provide a more comprehensive image editing and generation suite.
5. **Handling Larger Image Sizes**: Explore approaches for generating larger images without losing detail, perhaps through a multi-step generation process or by applying advanced upscaling algorithms.
6. **Exploring Multimodal Inputs**: Experiment with multimodal inputs such as combining text prompts with image inputs (e.g., style transfer), which could result in richer image generation with more diverse applications.
7. **Addressing Ethical Concerns**: Incorporate safeguards to prevent the generation of harmful, biased, or inappropriate content. This might include content filtering, better monitoring of input prompts, and more rigorous ethical guidelines.
   1. **Conclusion:**

The project on image generation using Stable Diffusion and Comfy UI has successfully demonstrated the potential of generative models in creative applications. By leveraging the capabilities of Stable Diffusion, the system offers a user-friendly interface

that empowers users to generate high-quality images based on textual descriptions. This contribution highlights the power of AI in content creation, enabling diverse creative outputs ranging from artistic representations to design mockups.

The project also demonstrates the integration of machine learning with user interfaces, making complex models accessible to a wider audience without requiring deep technical expertise. Although the system works well, there are areas for improvement, such as optimizing speed and expanding customization options.

Nevertheless, the impact of this work lies in its ability to democratize creative processes, offering tools that could be used in industries such as design, entertainment, marketing, and more.

By focusing on future developments, the project can evolve to meet the growing demands for more personalized, efficient, and ethically sound AI-driven tools for content creation.

**REFERENCES**

**[1] Sasirajan, G., Guhan, M., Mary, R., Maheswari, R., & Roselin, M. April 2023. Image generation with Stable Diffusion AI. IJARCCE,** **Vol. 12, Issue 5, Pgs. 643-647.**

**[2] Shetty, M., Sheikh, H., Sharma, P., Srivastava, K., & Gonsalves, J. February 2024. A comparative analysis of AI image generation models: Stable Diffusion, Dall-E and Dream by WOMBO. International Journal of Creative Research Thoughts (IJCRT), Volume 12, Issue 2, Pgs. c278-c281.**

**[3] Singh, A. April 2024. Stable Diffusion model: The past, present, and the future in India. International Journal of Scientific Development and Research (IJSDR), Volume 9 Issue 4, Pgs. 22-32.**