# Project Presentation

# "Gen-AI in Blender: AI-driven prompt-based image generation”

**MITU21BTCS0221:- Harkeerat Dhunda**
**MITU21BTCS0718:- Vilakshan**

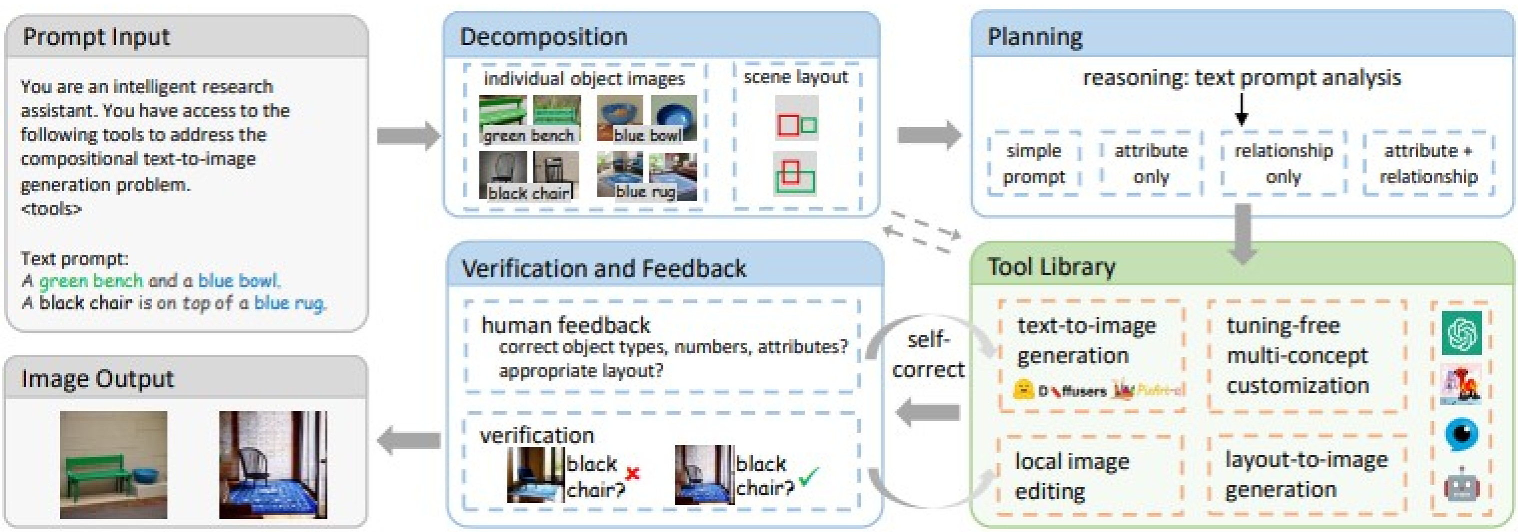
**Guided by**
**Prof. Sneha Singha**

# Outline

- Introduction
- Problem Statement
- Objectives
- Concepts & Methods
- Literature Review
- Identification of gaps & scope of work
- Implementation
- Reference
- Q&A

# 1. Introduction

AI picture generators, powered by trained neural networks, create lifelike graphics from text inputs while smoothly combining styles and themes. This breakthrough, based on Generative AI, uses large datasets to create contextually relevant artwork. Notable techniques include neural style transfer, GANs, and diffusion models, each with unique characteristics.



# 2. Problem Statement

Current limitations in Blender's integration for AI-driven image generation create a positive opportunity to enhance the user experience, enabling a smoother and more accessible incorporation of AI-generated content in 3D design projects.

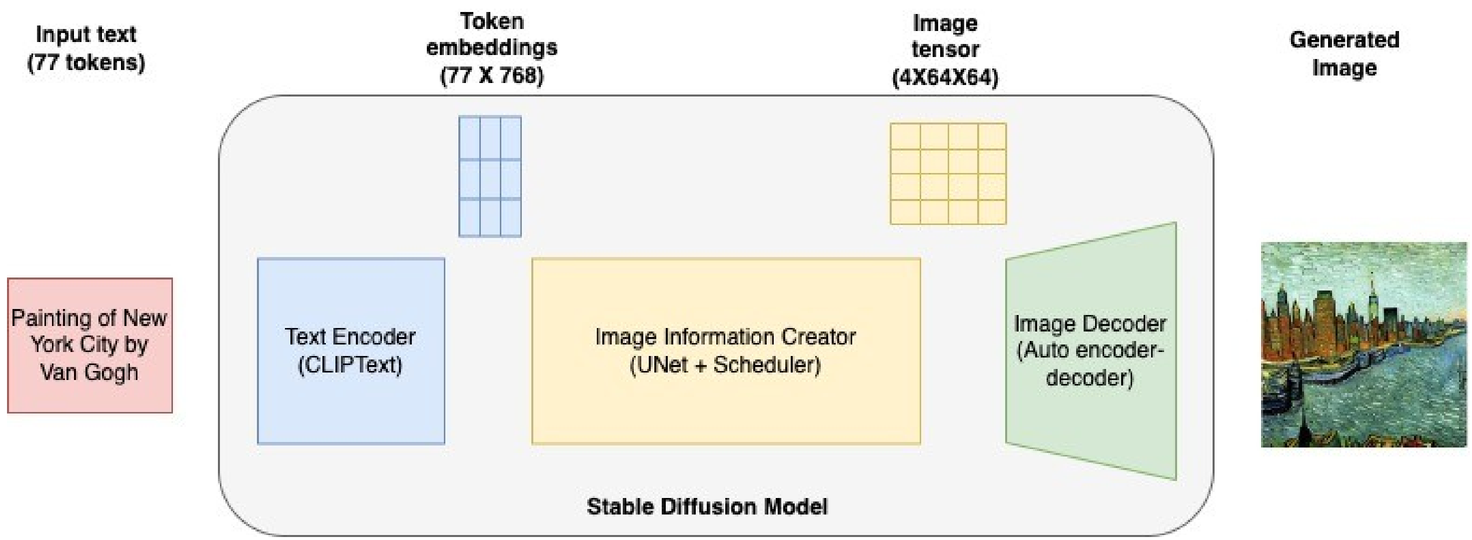
# 3. Objectives

- Develop an add-on for Blender that integrates Generative AI technology
- Create an easy-to-use interface for prompt based interactions with AI algorithms
- Encourage AI-generated content to be seamlessly integrated into 3D designs
- Utilize innovative GANs like StyleGAN
- Help artists create beautiful 3D designs with easy-to-follow instructions

# 4. Concepts and Methods

**1. Stable Diffusion**

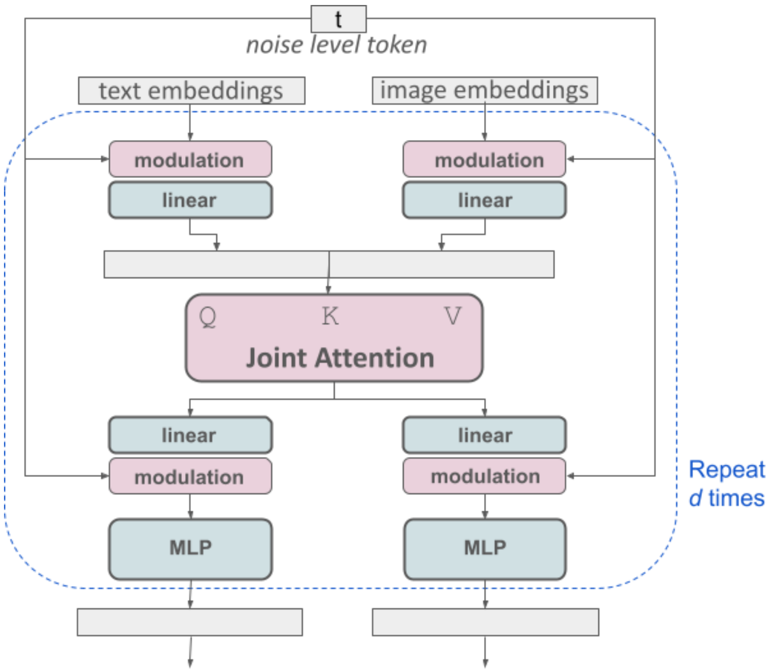
Stable Diffusion utilizes a combination of a variational autoencoder (VAE), U-Net, and optionally a text encoder. The VAE compresses the image into a smaller latent space, while Gaussian noise is applied iteratively to this representation during forward diffusion. The U-Net, built on a ResNet backbone, then denoises the output back to a latent representation, which is finally decoded by the VAE into the original pixel space.



# 4. Concepts and Methods

**2. Modified multimodal diffusion transformer: MMDiT**

For text-to-image generation, the chosen model has to take both modalities, text and images, into account. This is why we call this new architecture MMDiT, a reference to its ability to process multiple modalities. As in previous versions of Stable Diffusion, we use pretrained models to derive suitable text and image representations. Specifically, we use three different text embedders: CLIP models and T5 - to encode text representations, and an improved autoencoding model to encode image tokens.



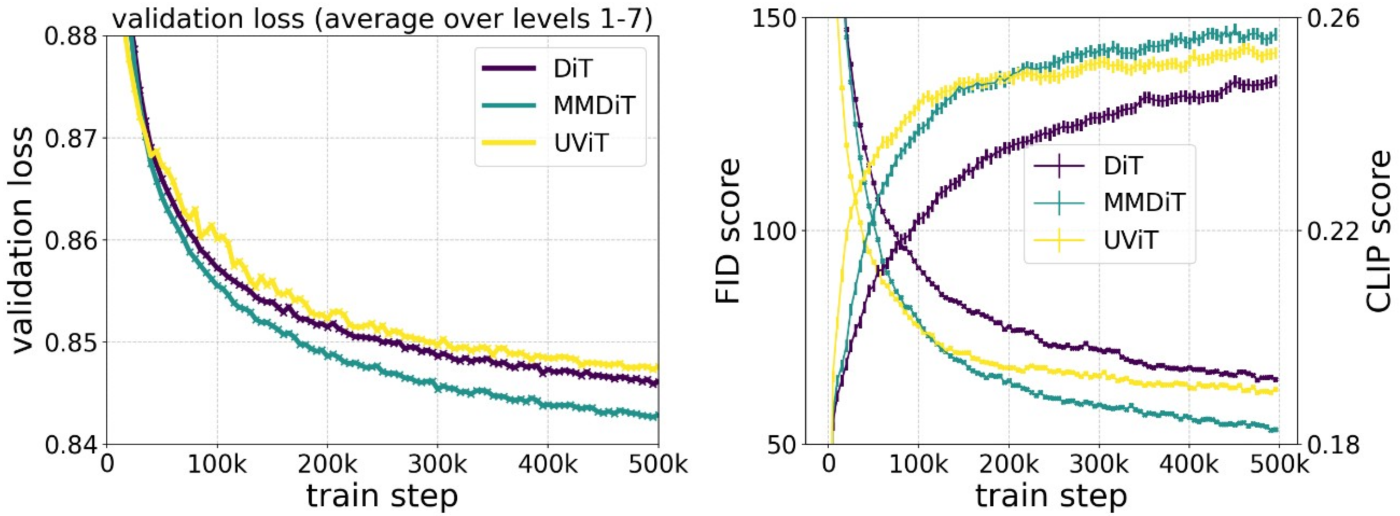


Prompt: Beautiful pixel art of a Wizard with hovering text "Achievement unlocked: Diffusion models can spell now"

# 4. Concepts and Methods

**3. Evaluation of Modified Multimodal diffusion transformer: MMDiT**

Validation loss measures how different the generated content is from the real content, with lower values indicating better performance. Visual fidelity score specifically evaluates how visually similar generated images are to real ones, with higher scores indicating better quality. Both metrics are crucial for assessing the performance of genAI models.



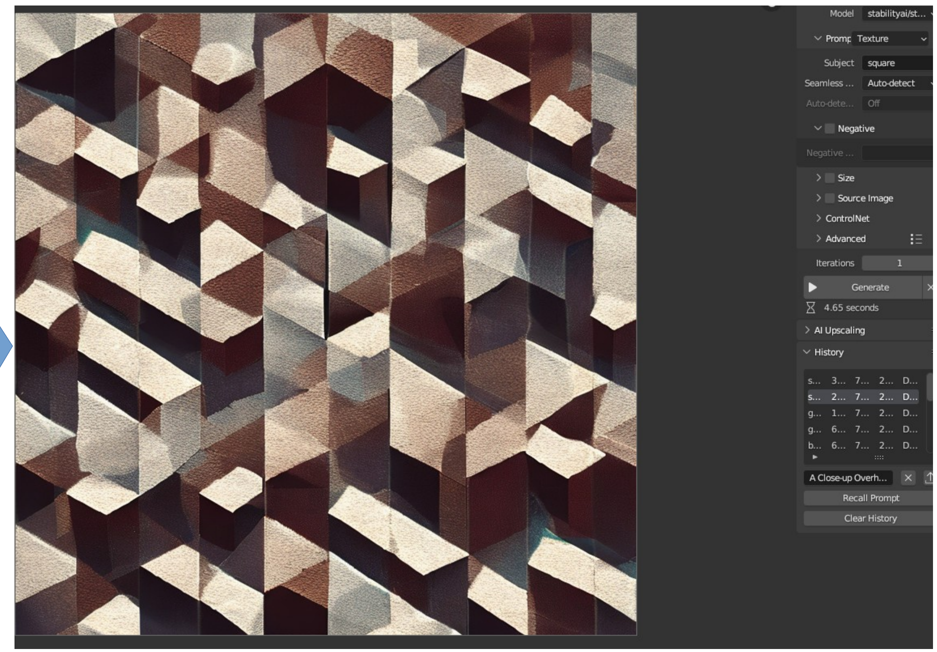

*Our novel MMDiT architecture outperforms established text-to-image backbones such as UViT (Hoogeboom et al, 2023) and DiT (Peebles & Xie, 2023), when measuring visual fidelity and text alignment over the course of training.*

# 5. Literature Survey

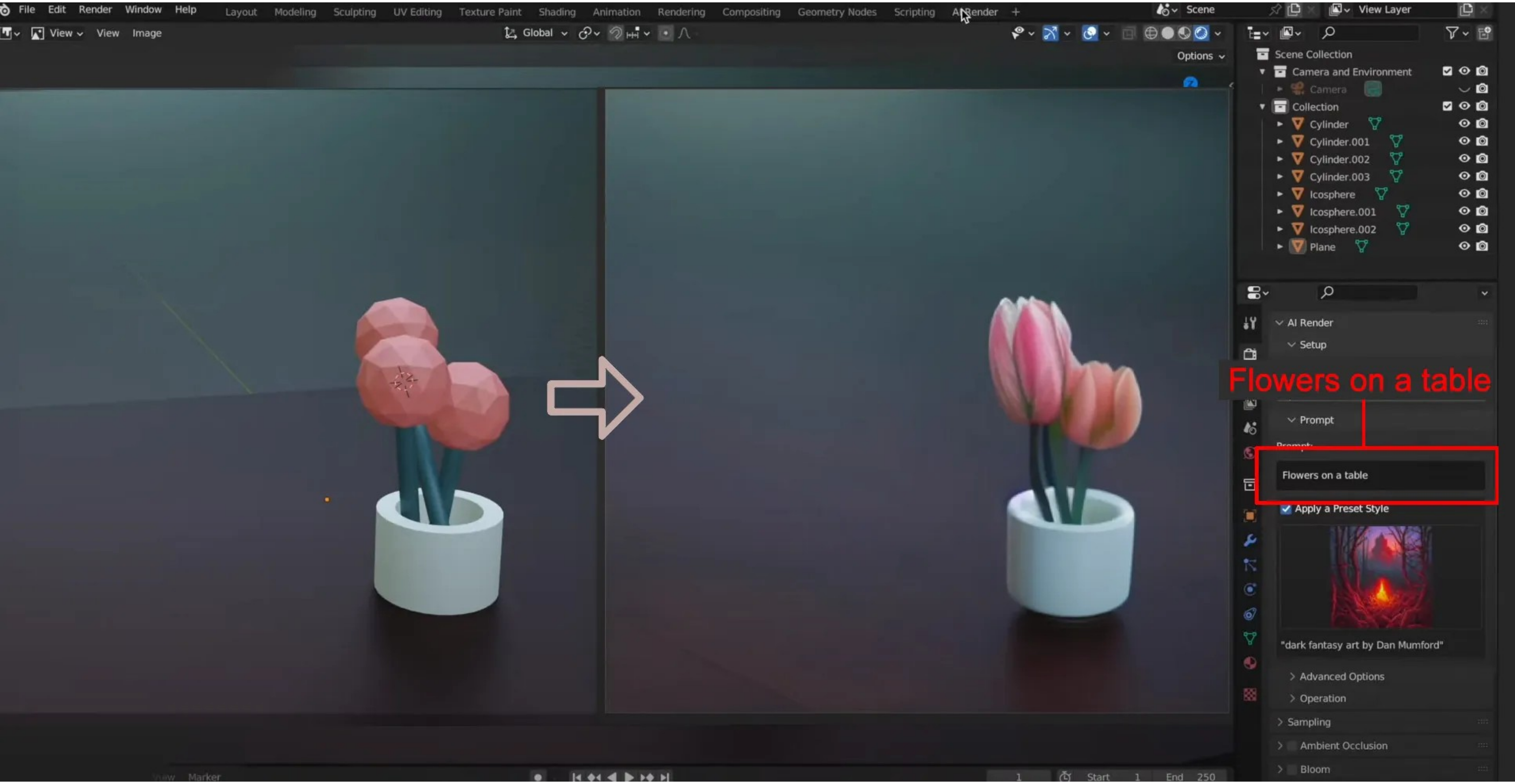
| Year | Method | Results | Drawbacks |
|---|---|---|---|
| 2023 | Utilization of Diffusion or GAN image generation methods (Stable Diffusion XL, GigaGAN). | GAN methods better for high-frequency details, with advantages in speed and precision. | Maintaining consistency across multiple images in 3D generation. |
| 2023 | Readily available generative AI on artistic practices | Focus on prompt engineering and design in the artistic process | Ready-made models might constrain the level of personalization |
| 2023 | Taxonomy of Prompt Modifiers | Six types identified: subject terms, image prompts, style modifiers, quality boosters, repeating terms, and magic terms. | Limited overall working time to about 2 hours per day, depending on computation |
| 2022 | Translating Raw Descriptions into Images by Prompt-based Cross-Modal Generation | Proposed PCM-Frame leveraging CLIP and StyleGAN pre-trained models. | Challenges in handling extremely abstract or subjective terms |
| 2021 | Generative Adversarial Networks for Image and Video Synthesis | . Deep Boltzmann Machines (DBMs) use MCMC sampling, but scalability is limited. Variational AutoEncoders (VAEs) excel in learning latent representations but might produce slightly blurry images. | Evaluating and comparing GAN models is challenging due to instability. |

# 6. Implementation

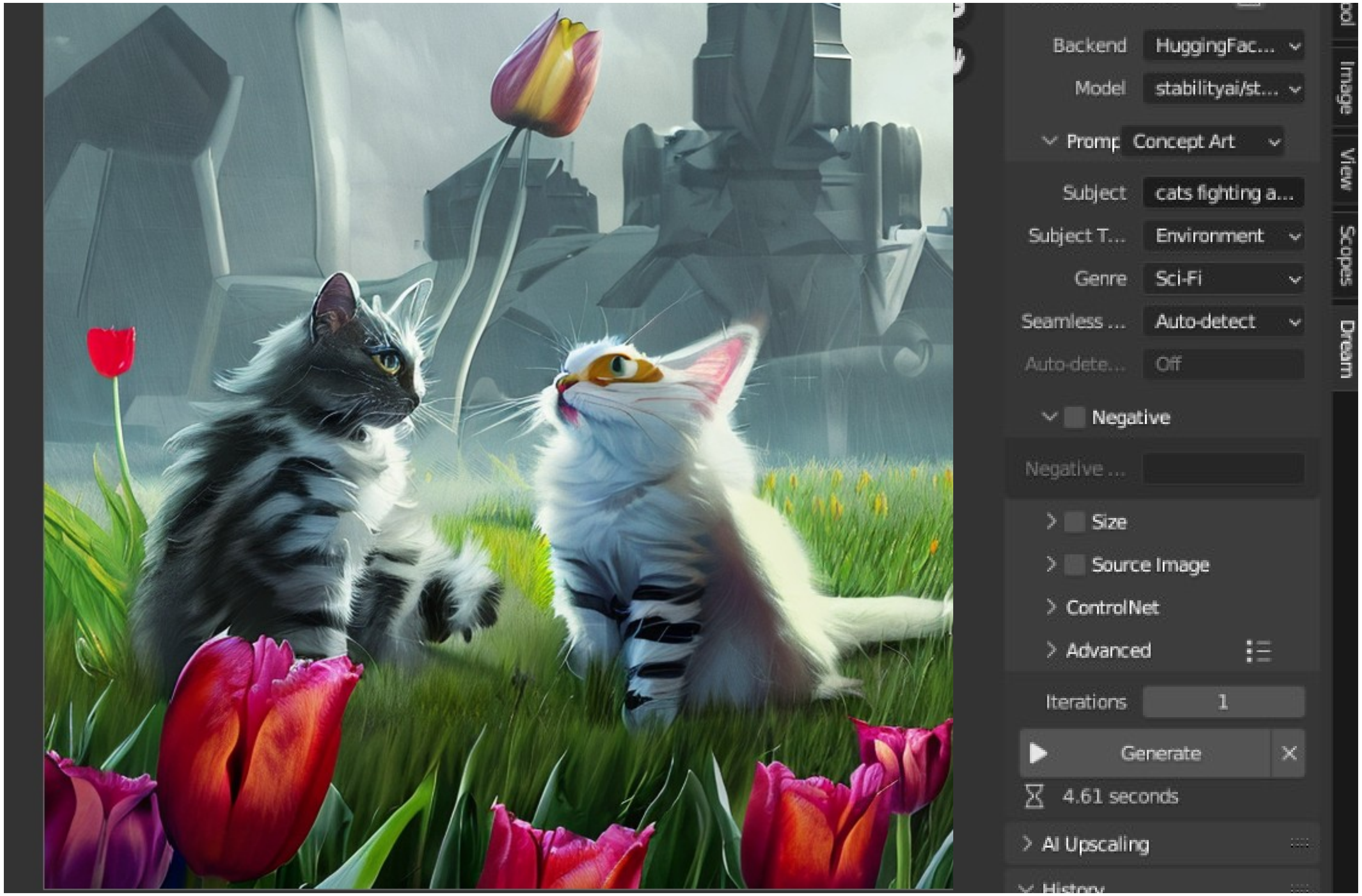
```python
import bpy
import numpy as np

def stable_diffusion(mesh_object, iterations=10, dt=0.1, alpha=0.5):
    mesh = mesh_object.data
    vertices = np.array([v.co for v in mesh.vertices])

    for _ in range(iterations):
        laplacian = np.zeros_like(vertices)
        for edge in mesh.edges:
            v1_idx, v2_idx = edge.vertices
            v1, v2 = vertices[v1_idx], vertices[v2_idx]
            laplacian[v1_idx] += (v2 - v1)
            laplacian[v2_idx] += (v1 - v2)

        vertices += dt * alpha * laplacian

        for i, v in enumerate(mesh.vertices):
            v.co = vertices[i]

        mesh.update()

bpy.ops.object.select_all(action='DESELECT')
bpy.ops.object.select_by_type(type='MESH')
bpy.ops.object.delete()

verts = [(0, 0, 0), (1, 0, 0), (1, 1, 0), (0, 1, 0)]
faces = [(0, 1, 2, 3)]
mesh = bpy.data.meshes.new(name="New Mesh")
mesh.from_pydata(verts, [], faces)
mesh.update()

obj = bpy.data.objects.new(name="New Object", object_data=mesh)
bpy.context.collection.objects.link(obj)

stable_diffusion(obj, iterations=100, dt=0.05, alpha=0.1)
```



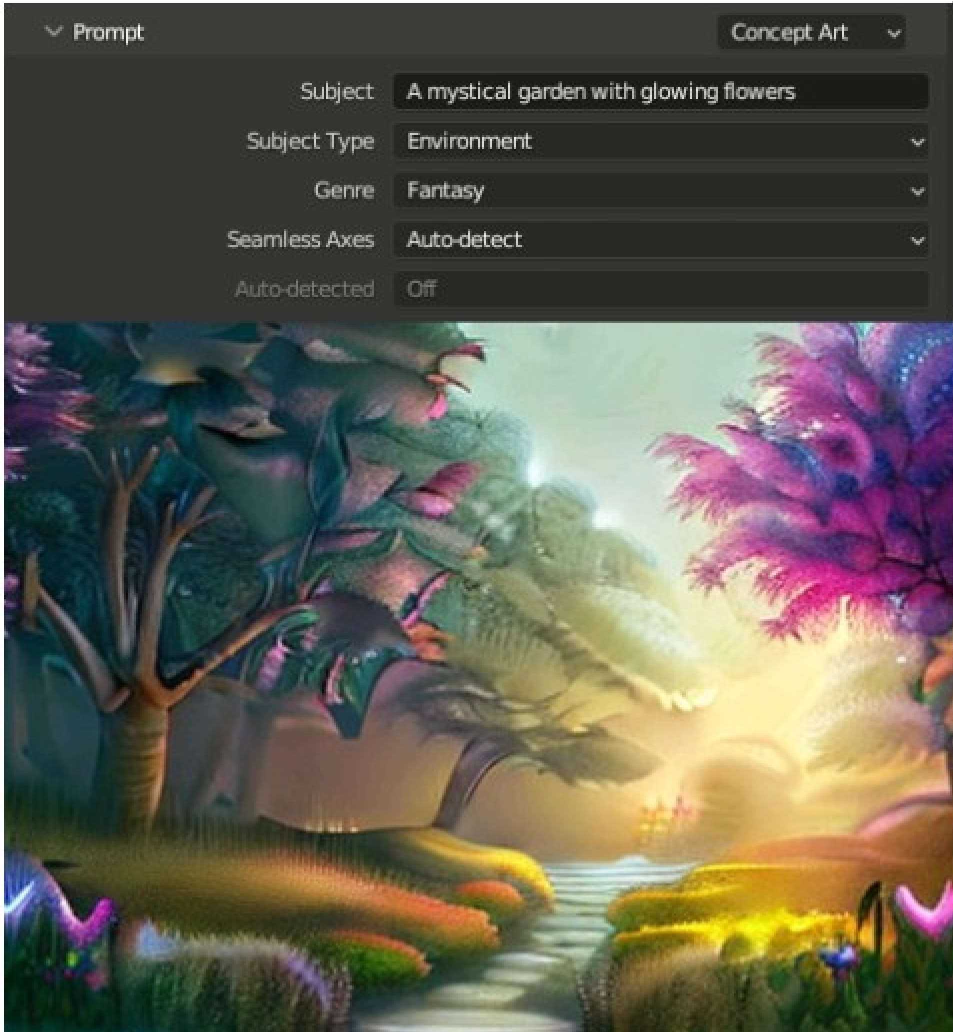
# 6. Implementation

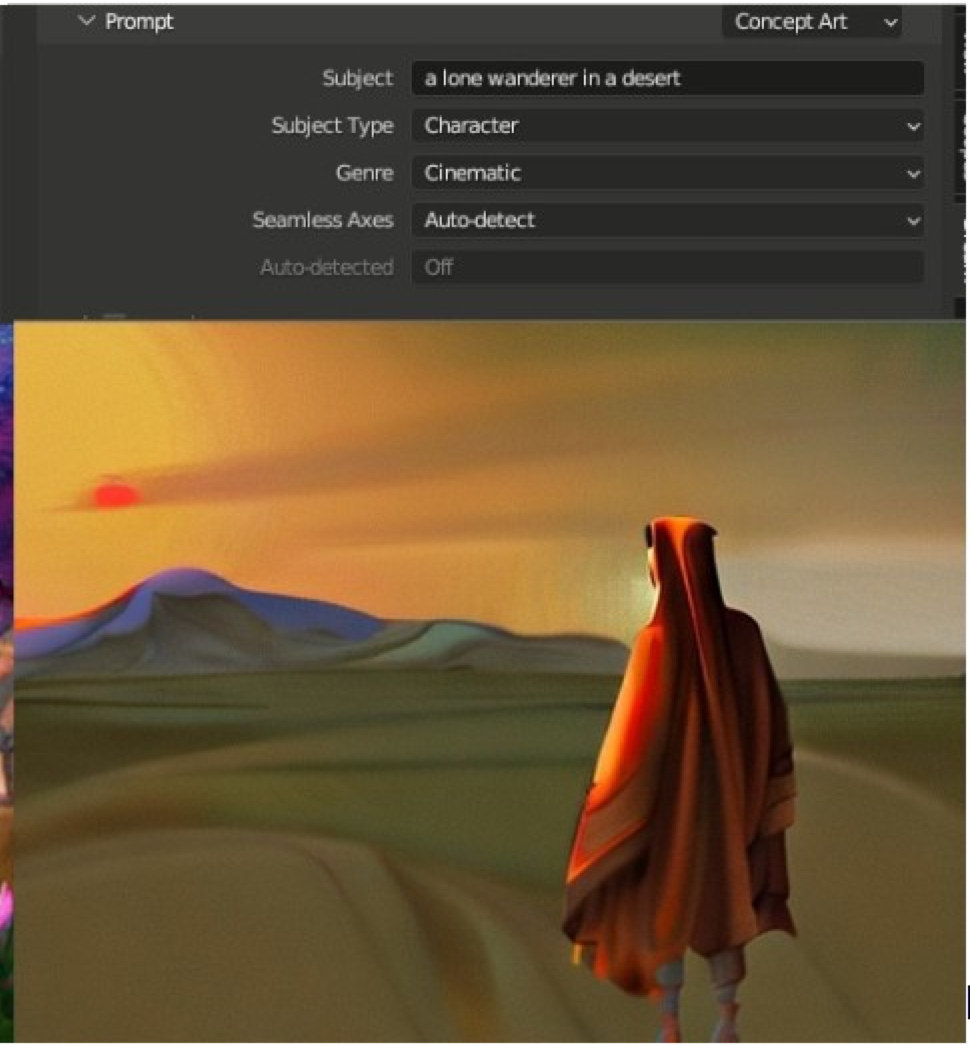


# 6. Implementation



# 7. Results


Prompt
Concept Art
Subject
A mystical garden with glowing flowers
Subject Type
Environment
Genre
Fantasy
Seamless Axes
Auto-detect
Auto-detected
Off


Prompt
Concept Art
Subject
a lone wanderer in a desert
Subject Type
Character
Genre
Cinematic
Seamless Axes
Auto-detect
Auto-detected
Off

# 7. Results

**1. Visual Quality:**

High visual quality involves sharp details of the images with vivid colors and realistic textures. The glowing flowers in the mystic garden and the lone wanderer in the desert are both clear, and in concordance with the input description. It shows fine-grained detail, where the AI model detects complex patterns on the glowing flowers and a large expanse in the desert

**2. Artistic Interpretation:**

The palette, the lighting effects, and equally the composition of the scenes bring the emotion in such a way that the audience gets invited into the world filled with wonder and curiosity. The otherworldly beauty of the magical garden and the feeling of loneliness over self in the desert landscape show how versatile and imaginative the AI model could be in trying to visualize these two rather diverse types of visual stories.

# 7. Results

**3. Enhanced Model's Ability:**

- Consistently interpret and apply diverse artistic styles.
- Explored neural architecture improvements for style consistency
- Researched state-of-the-art style.

**4. Expanded Prompt Diversity:**

- Increased the diversity of prompts for generating a wider range of images.
- Analyzed user behavior to understand prompt preferences
- Implemented a prompt suggestion system.

# 8. Future Work

## FreeU: Free Lunch in Diffusion U-Net

Chenyang Si, Ziqi Huang, Yuming Jiang, Ziwei Liu
S-Lab, Nanyang Technological University
Paper | Project Page | Video | Demo

Demo 🤗 Hugging Face

**We propose FreeU, a method that substantially improves diffusion model sample quality at no costs: no training, no additional parameter introduced, and no increase in memory or sampling time.**


📖 For more visual results, go checkout our project page

# 9. Conclusion

The integration of Generative AI in Blender represents a paradigm shift in content creation, empowering users to push the boundaries of their creativity and unlock new artistic horizons. With its vast potential and versatility, GenAI in Blender is poised to revolutionize the way we approach digital content creation in the years to come.

**Optimize Image Resolution**

Further optimize algorithms for generating even higher resolution images.
Explore distributed computing for parallel image generation.

# References


**1. Hutson, J., & Robertson, B. (2023).** Exploring the Educational Potential of AI Generative Art in 3D Design Fundamentals: A Case Study on Prompt Engineering and Creative Workflows. Global Journal of Human-Social Science: Arts & Humanities - Psychology. https://doi.org/10.1145/3503161.3547790

**2. Oppenlaender, J. (2023).** A Taxonomy of Prompt Modifiers for Text-To-Image Generation. https://digitalcommons.lindenwood.edu/faculty-research-papers/485/

**3. Kang, S., et al. (2023).** Scaling Up GANs for Text-to-Image Synthesis. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content/CVPR2023/html/Kang_Scaling_Up_GANs_for_Text-to-Image_Synthesis_CVPR_2023_paper.html

**4. Sagescribe, A. (n.d.).** Generative AI for Image Generation: SOTA, Common Methods. Medium. https://medium.com/@aisagescribe/generative-ai-for-image-generation-sota-common-methods-bea0a70c9b81

# Thank You