
Lil Gen
Text To Image
Generator
Supervised By-Sa Phyo Thu Htet

Group Members

Min Phone Thit

Speaker-1

Ye Bhone Lin

Speaker-2

Agenda

01 Problem Statement

02 How StableDiffusion Works

03 Data Preprocessing

04 Experiments

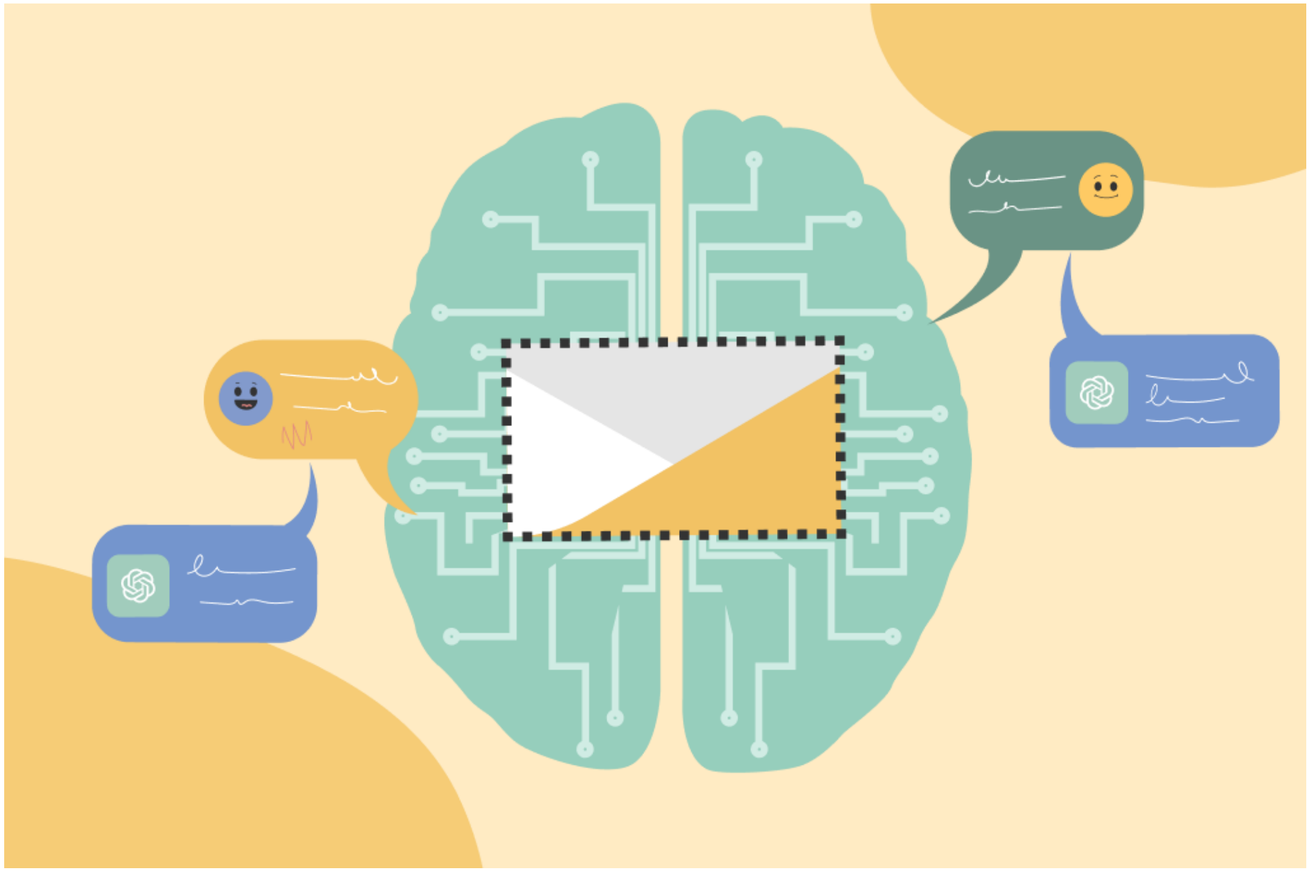
05 Demonstration

06 Challenges

07 Future Works

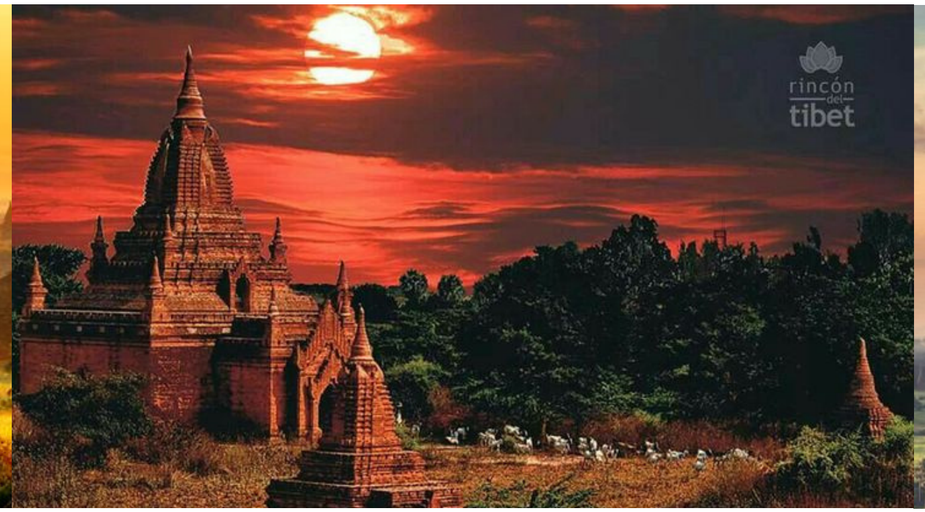
08 Ai Ethics

09 References



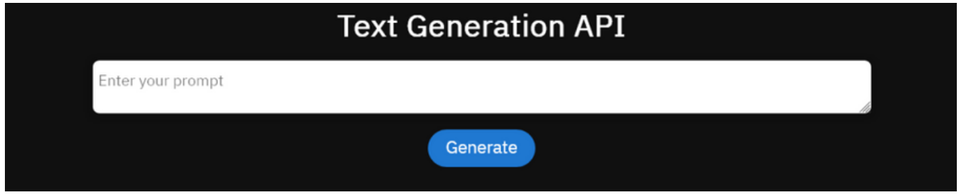
About Our Project

- Fine Tuning the stable diffusion model with famous places in Myanamar


What is AI Text To Image Generator

AI text-to-image generators basically use user’s written description to create an image that matches the prompt.



These are popular ai text to image generators,

Mid Journey

Blue Willow

Canvas AI

Problem Statement

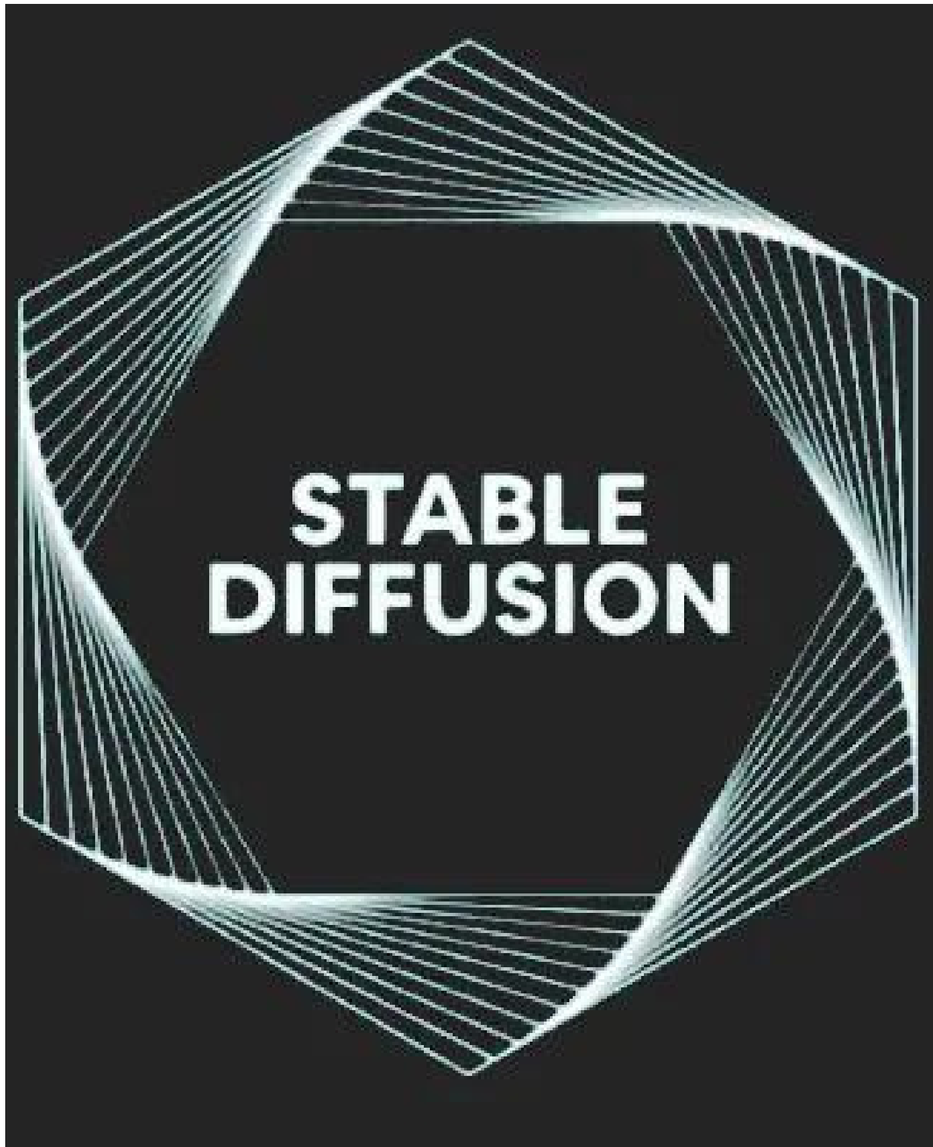
When we prompted the stable diffusion model (v-1.5) to generate an image of Bagan, it produced an image depicting a pagoda from Thailand.

Hence, our decision was to fine-tune the current stable diffusion model using a multitude of Bagan photos in order to attain a clearer outcome.

prompt - astonishing view of <Bagan> Pagoda

Stable Diffusion

- deep learning model that uses diffusion processes to generate high-quality artwork from textual input
- it utilizes a novel training method called "stable training" to ensure that the generated images are of high quality and consistent with the textual input
- stable diffusion algorithm is capable of producing a wide range of artistic styles, including photorealistic portraits, landscapes, and abstract art

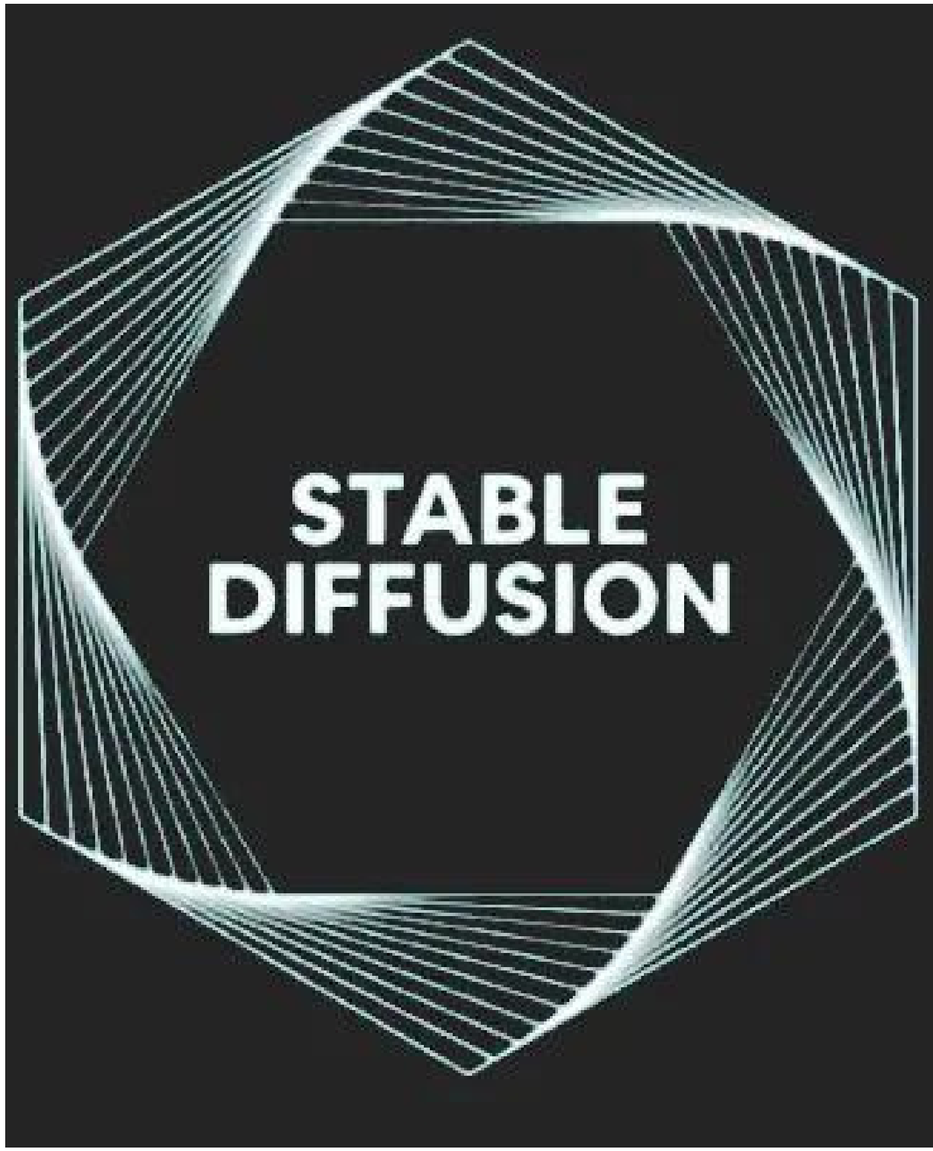


How Stable Diffusion Works

01

Text Interpretation

- User inputs a prompt in natural language.
- Stable Diffusion interprets and comprehends the text request
- Pertinent information is extracted from the text to generate the intended image

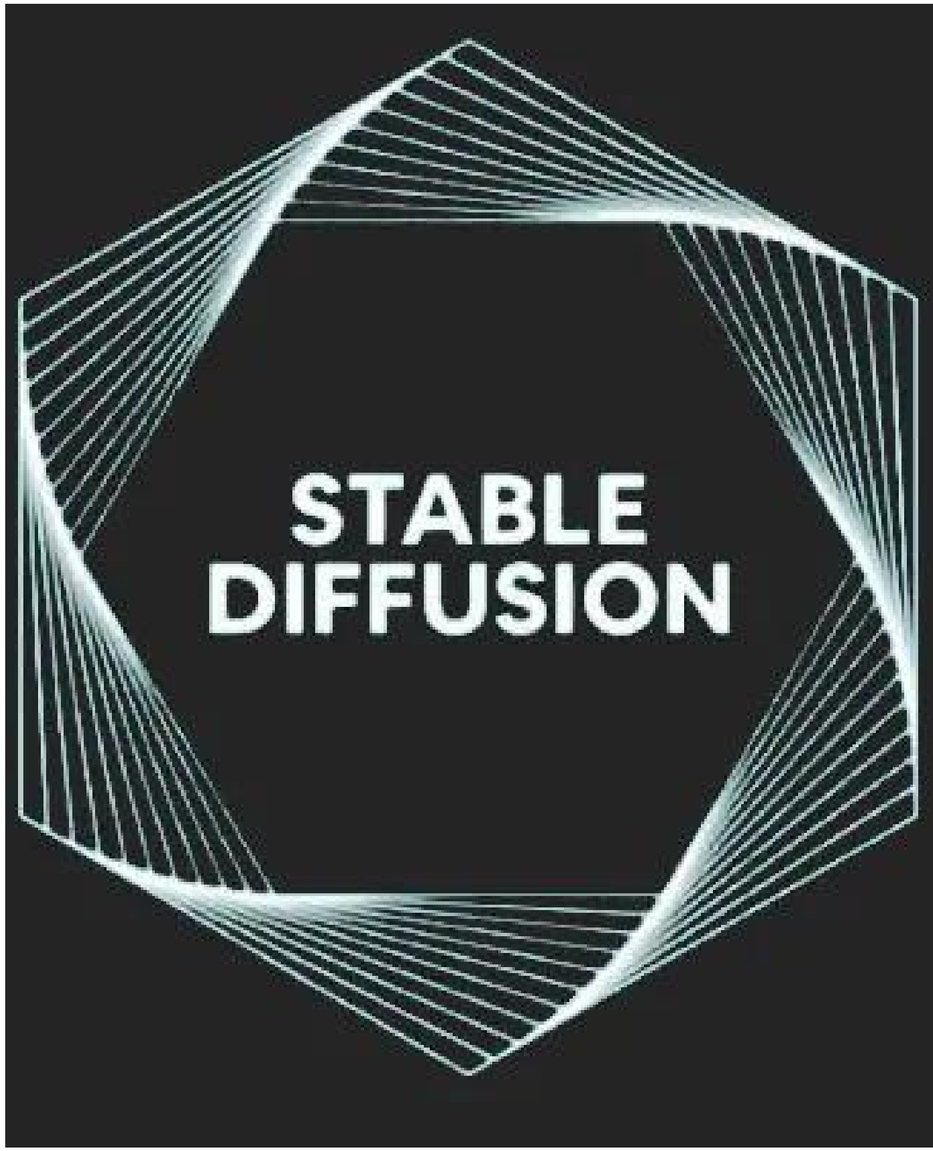


How Stable Diffusion Works

02

Employs a Diffusion Model

- Stable Diffusion employs a diffusion model specifically trained for image generation.
- The model produces a noisy and blur image.
- The model then eliminates Gaussian noise from initially blurry images.

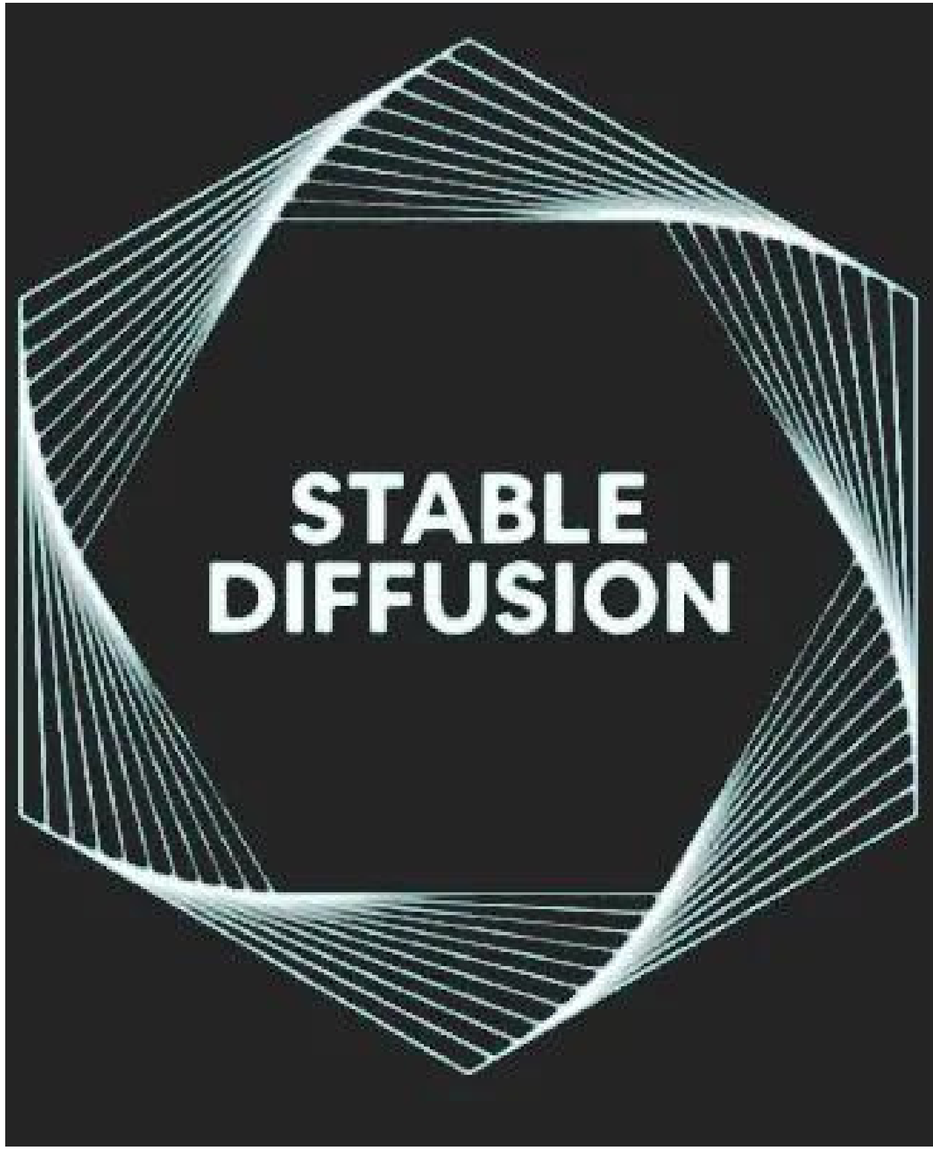


How Stable Diffusion Works

03

Diffusion Process

- Stable Diffusion operates iteratively until the diffusion process is done
- At each iteration, the algorithm calculates the diffusion coefficient from local image features.
- Adjusts the colors or intensities of pixels based on specific rules or calculations.

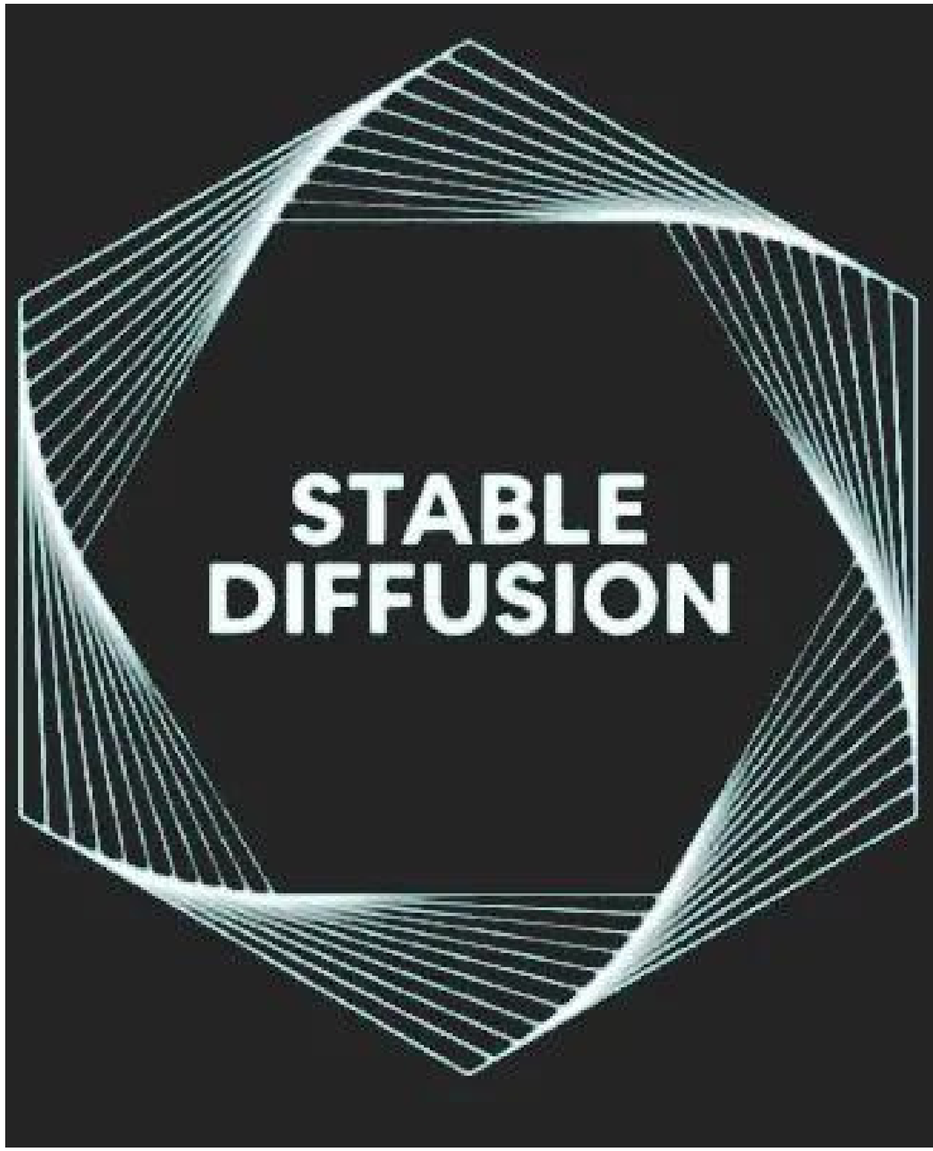


How Stable Diffusion Works

04

Image Generation

- The model uses the textual information provided by the user as well as its training knowledge to generate the image.
- After understanding the text description and applying the diffusion model, Stable Diffusion creates an image.



Data Preprocessing

01 Teach the model

02 The Autoencoder (VAE)

03 UNET

04 Text-Encoder

Teach The Model

There are two types of styles to teach the model.

Object

enables you to teach the model a new object to be used

Style

allows you to teach the model a new style one can use

The Autoencoder (VAE)

A latent space -> a mathematical space which maps what a neural network has learnt from training images.

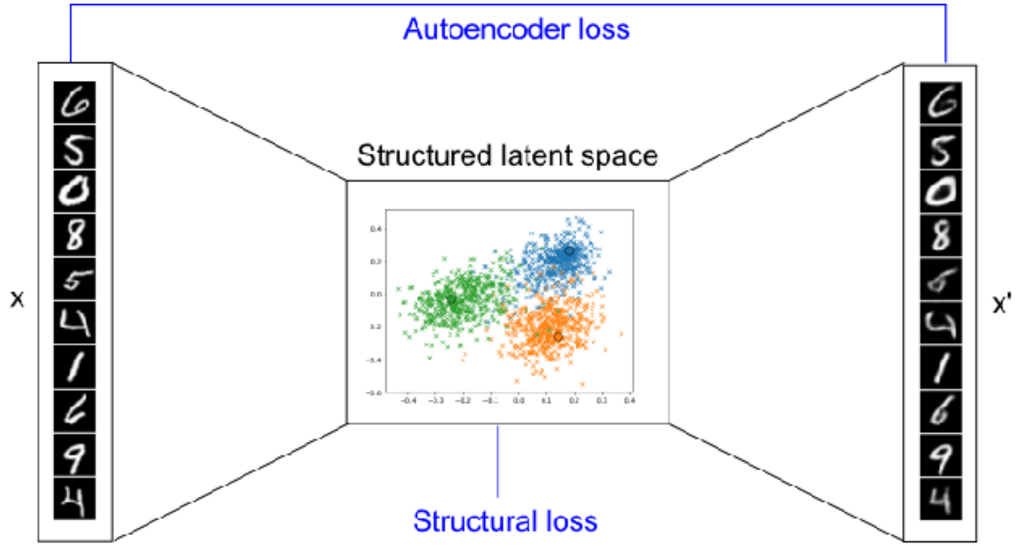
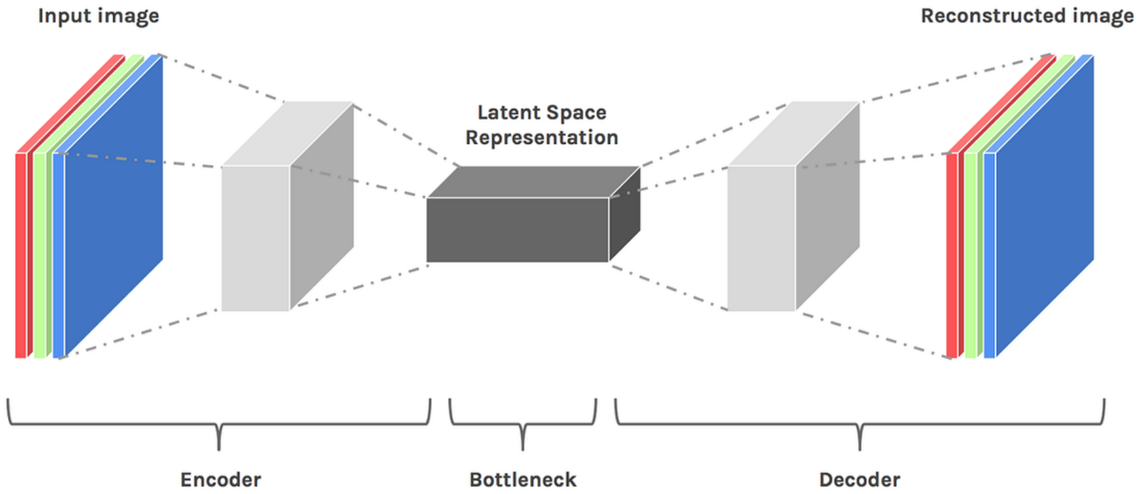


Figure 2. Our Structuring AutoEncoder (SAE) projects data into a

The encoder is used to convert the image into a low dimensional latent representation, which will serve as the input to the U-Net model

---

The decoder, conversely, transforms the latent representation back into an image.



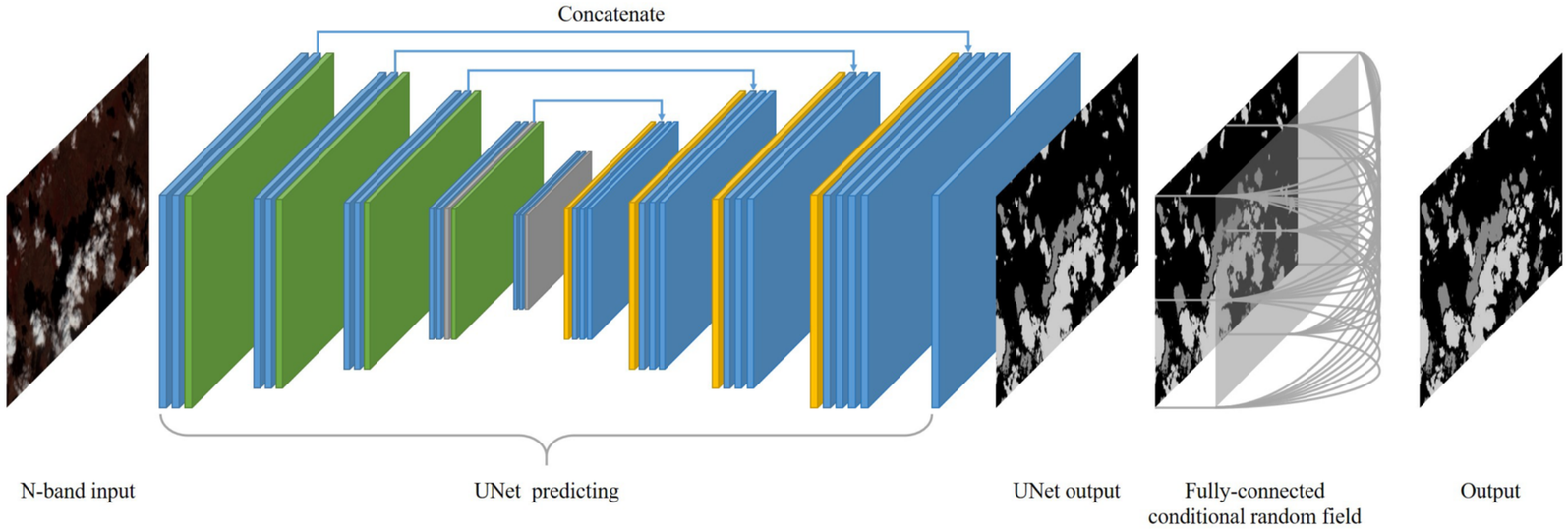
Resize the input images to 512x512 image

1500x1000

512x512

UNET

To interpolate those patterns and remove noise at the same time.

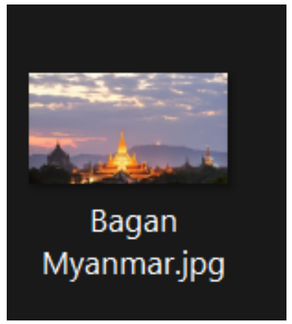


Noise Input Image

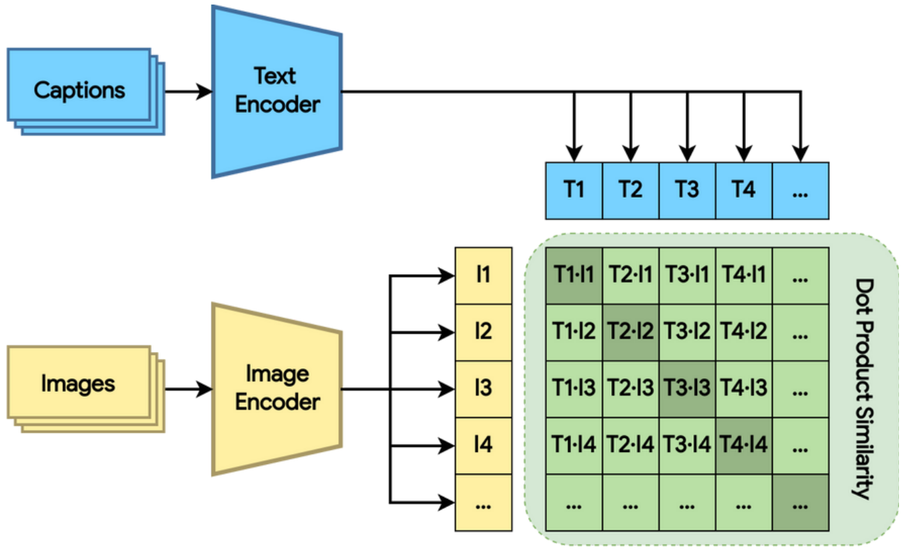
Denoise Image

Text-Encoder

Text-Encoder creates embeddings corresponding to the input text.

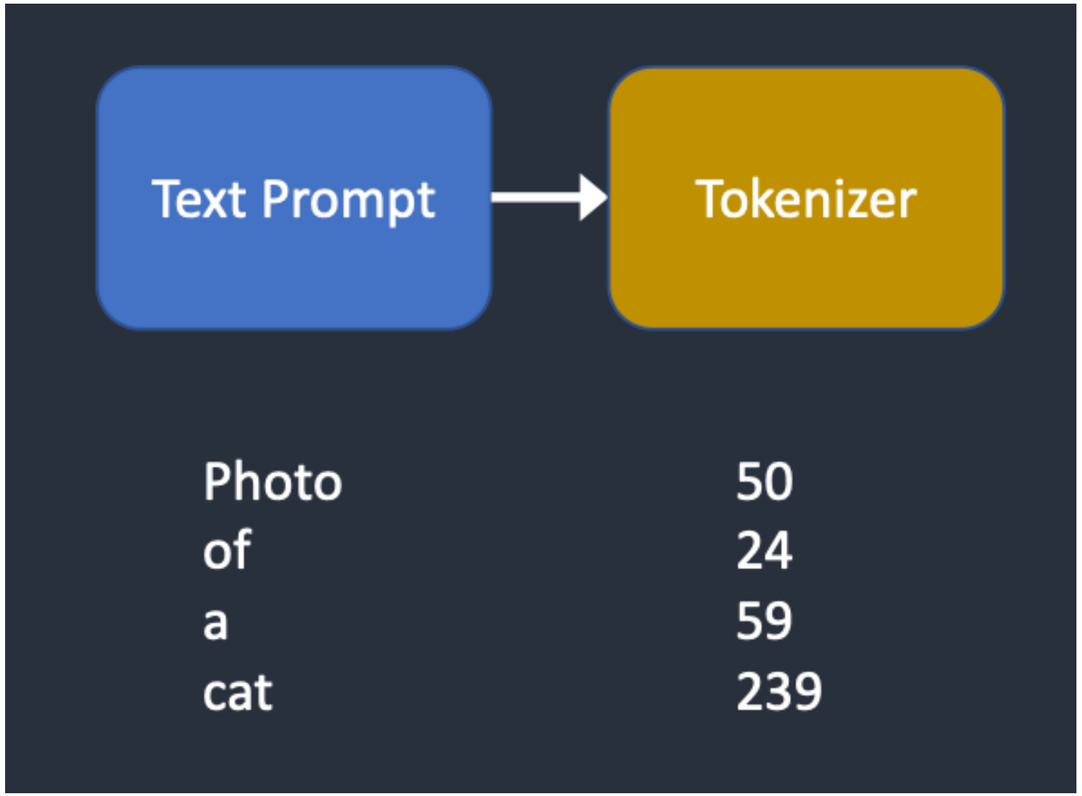


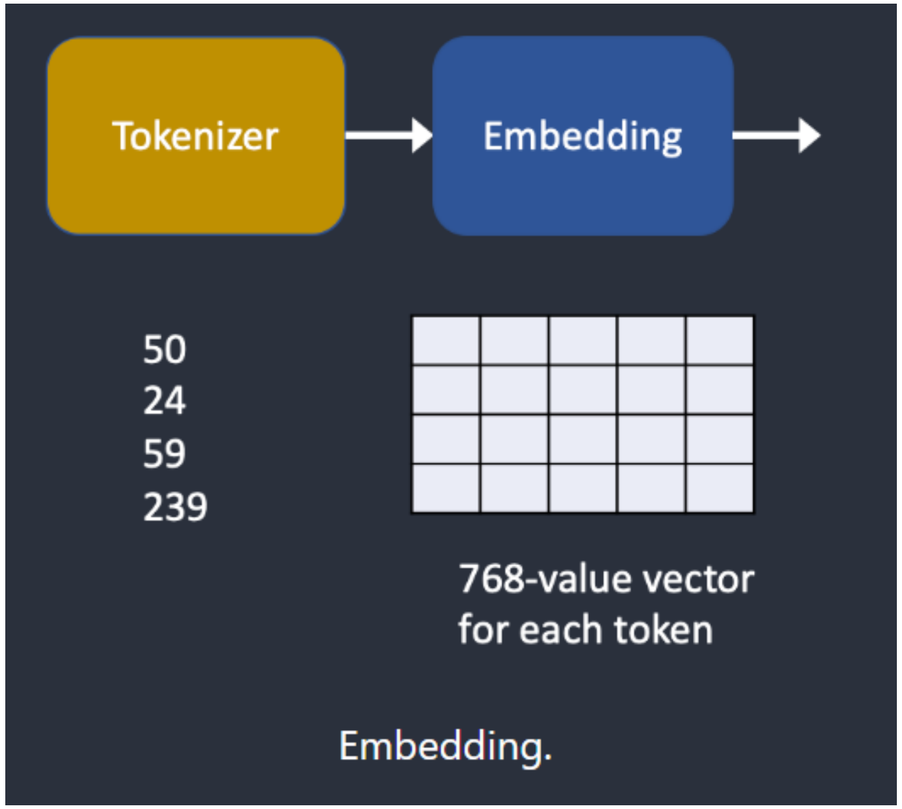
It takes texts from the image and generates numerical representations or embeddings of that text.



https://medium.com/@onkarmishra/stable-diffusion-explained-1f101284484d#:~:text=The%20text%20encoder%20is%20usually,corresponding%20to%20the%20input%20text.

Tokenizes a text prompt into a sequence of tokens.





Eg. "Photo of a cat" into an embedding space that can be understood by the U-Net.

https://stable-diffusion-art.com/how-stable-diffusion-work/#:~:text=Stable%20Diffusion%20v1%20uses%20CLIP's,it%20has%20seen%20during%20training.

Experiments

Our experiments aim to train a text encoder in a novel way to enhance its ability to generate meaningful embeddings.

We adopted a multi-step approach involving text encoding, noise injection, and gradient optimization.

- Stable diffusion v1.1,v1.2,v1.3,v1.4,v1.5
- We used stable diffusion v1.5 model to train with 99 bagan pictures.

Hardware	Software
We utilized NVIDIA GPUs for accelerated training.	PyTorch and Accelerate for efficient GPU utilization.

Evaluation

Text-Encoder

Simple transformer-based encoder that maps a sequence of input tokens to a sequence of latent text-embeddings.

Human Evaluation

Human evaluation involves taking surveys from various people and evaluating their scores based on their feedback.

For text-encoder we used CLIP Score

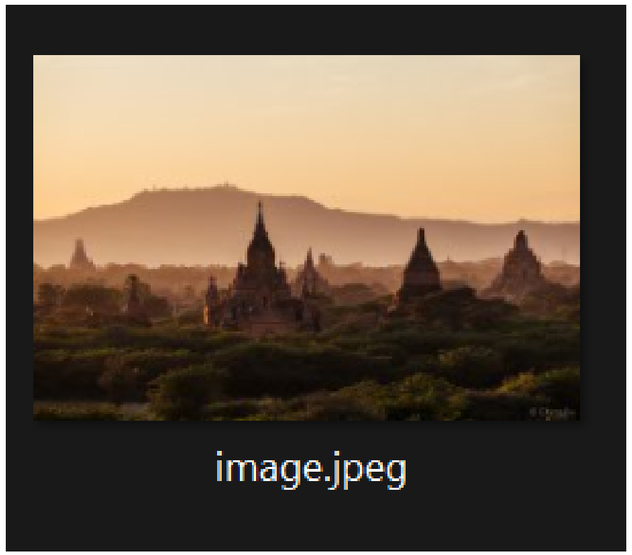
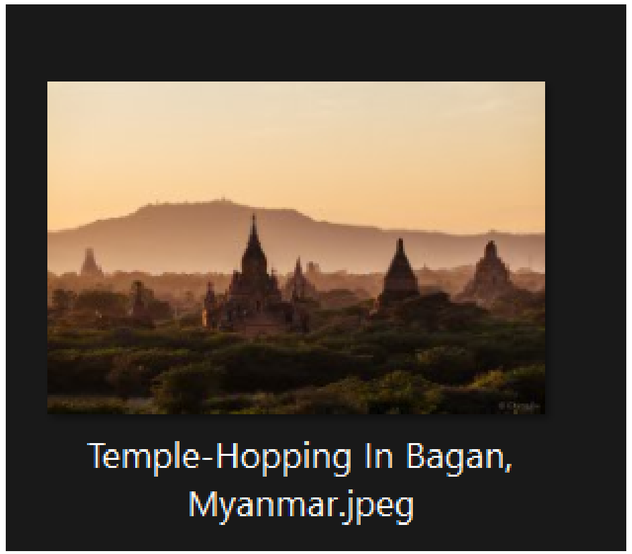


image.jpeg

Stable Diffusion V1.5 -> 20.93

Stable Diffusion V1.4 -> 20.75



Temple-Hopping In Bagan, Myanmar.jpeg

Stable Diffusion V1.5 -> 35.85

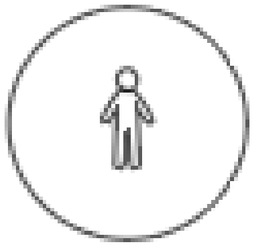
Stable Diffusion V1.4 -> 35.50

Human Evaluation

Users

4 females

6 males



Made a survey for rating photos

- 3 Ai Generated Bagan Photos(hyperparameter tuning)
- 1 Original Bagan Photo from Google

Hyperparameter Tuning

Hyperparameter	Learning Rate	Max_train_steps	Batch_size
Hyperparameter1	5e-4	400	4
Hyperparameter2	3e-6	500	5
Hyperparameter3	6e-6	600	4

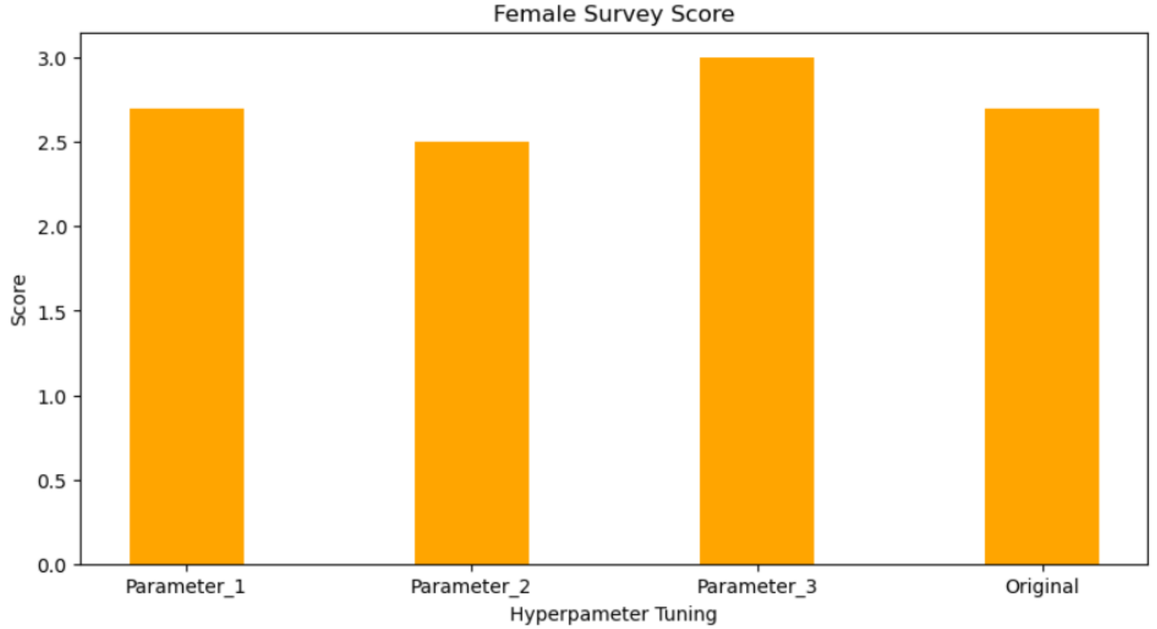
Human Evaluation

Please Rate these photos.

- ◯ 1. Very Satisfied
- ◯ 2. Satisfied
- ◯ 3. Neutral
- ◯ 4. Dissatisfied
- ◯ 5. Very Dissatisfied

https://docs.google.com/forms/d/e/1FAIpQLScI6FFiiFhKD4JlXFMk6C6Pi3jByjM078KUMVytPtssKdHojg/viewform?vc=0&c=0&w=1&flr=0

Experiments & Evaluation



Males mostly like parameter 2 generated photos

Females mostly like parameter 3 generated photos

Survey Results

Hyperparameter	Scores
Parameter 1	4.4
Parameter 2	4.3
Parameter 3	4.2
Original	3.9

Parameter-1

Demonstration

In this demo, we used Parameter1.
Let’s see how it works.

Demonstration Link

https://github.com/Billy1437/Lil-Gen-Text-To-Image-Generator.git

Challenges

01 Our demonstration lasts only 72 hours.

02 Photos of low quality.

03 We encountered insufficient GPUs memories.

Future Works

01 We will be deploying API for user-friendliness.

02 We are going to start training with photos of Mandalay and much more.

Ai Ethics

01

Privacy and Security

User text descriptions are stored securely. The platform adheres to strict data privacy regulations and only retains information necessary for improving the service.

02

Accountability

The system is designed to maintain strict adherence to content guidelines, thereby preventing users from incorporating inappropriate or offensive material as prompts within the model.

03

Transparency

Our model exclusively produces high-quality Bagan photos, while others are of lower quality.

References

01 Wikipedia (2022). Stable Diffusion. Retrieved From: https://en.wikipedia.org/wiki/Stable_Diffusion

02 Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. Retrieved From: https://arxiv.org/abs/2112.10752

03 Naomi Brown (2022). What is Stable Diffusion and How to Use it. Retrieved From: https://www.fotor.com/blog/what-is-stable-diffusion

Q&A

Thank You

for watching!