Generative AI

Generative ai is a subfield of Artificial Intelligence algorithms and models that are used to generate new data, such as images, text, audio, or even video, based on patterns learned from a training dataset. These models can create new content that resembles the input data.

Generative AI models include generative adversarial networks (GANs), variational autoencoders (VAEs), and autoregressive models, among others. These models have applications in various fields, including image generation, text generation,Text to video generation, and more.

Image Generation

Image generation refers to the process of creating new images using generative AI models. These models are trained on large datasets of images and learn to generate new, visually plausible images that resemble the training data.

There are several approaches to image generation from text prompt, including:

GANs(Generative Adverbial Networks): GANs consist of two neural networks, a generator and a discriminator, which are trained simultaneously. The generator creates fake images, while the discriminator tries to differentiate between real and fake images. Through this adversarial training process the generator learns to create increasingly realistic images through iteration process (DCGAN, StyleGAN, etc.).

VAEs(Variational Auto Encoders): VAEs are probabilistic models that learn a latent representation of images. They consist of an encoder that maps images to a latent space and a decoder that reconstructs images from the latent space. By sampling from the latent space, VAEs can generate new images (Vanilla VAE, DALL-E, etc.).

Autoregressive Models: Autoregressive models, such as PixelCNN and PixelRNN, generate images pixel by pixel. They model the conditional probability of each pixel given the previous pixels, allowing them to generate high-quality images with fine details. However, autoregressive models can be slow to generate images compared to other approaches (PixelCNN).

Diffusion Models: Diffusion models generate images by iteratively adding noise to an image and then removing the noise to generate the final image. They have been shown to generate high-quality images with realistic details and textures (Stable Diffusion, LoRA, etc).

Text to Image Generation models

We are using pre trained model which are deployed on the HuggingFace API keys. Initially I worked with the popular models for test to image generation models.

Stable-diffusion-2:

Stable diffusion model 2 is the refined model of the initial stable diffusion model released by stabilityai. The most common thing in every model is the inference steps or the image generation steps which decides the quality of the image generated with the given prompt. Int he stability diffusion model we were asked to create the image of the dog drinking coco-cola wearing green goggles. With different generation steps it preformed differently. This model took nearly 5 -6 steps to generate the image exactly referred by the prompt. You can change the output size of the image produced by the model and the number of iterations it have to take as input.

Dreamlike - diffusion V1:

Dreamlike Diffusion is a text-to-image diffuser model by Dreamlike Art, available through the Hugging Face model hub at dreamlike-art/dreamlike-diffusion-1.0. This model uses stable diffusion pipeline. The images generated by this model took less amount of time with 50 inference steps and more complex images are also generated by this model prove its efficiency in text to image generation.

Prompt: dreamlikeart, a grungy woman with rainbow hair, travelling between dimensions, dynamic pose, happy, soft eyes and narrow chin, extreme bokeh, dainty figure, long hair straight down, torn kawaii shirt and baggy jeans.

prompt: dog drinking coco cola with green colour goggles

Dream Shaper V 7:

Dream shaper v7 is a Stable Diffusion model that has been fine-tuned on runwayml/stable-diffusion-v1-5. The experiments are taken forward to check the quality of images produced by the number of inference steps.

Inference step = 1

Inference step = 3

Inference steps =10

Inference steps = 20

DALL-E 1.1 :

DALL-E 2 is a significantly more powerful version of the original model, and it can generate images that are more realistic, detailed, and varied. Dalle produced more realistic and better images in less time than the other models.

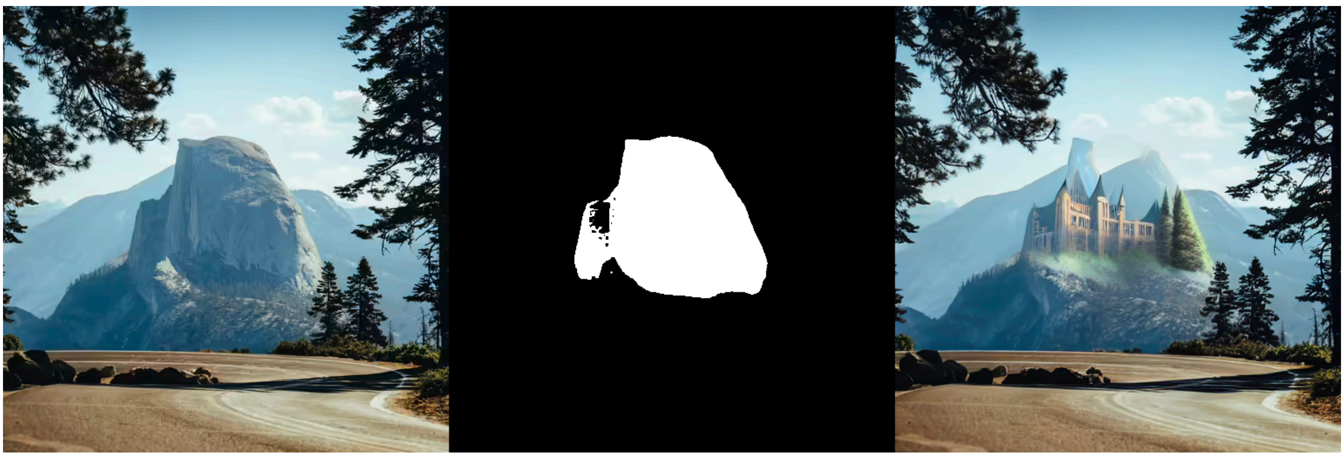
LCM - LoRA :

Latent Consistency Models (LCMs) have achieved impressive performance in accelerating text-to-image generative tasks, producing high-quality images with minimal inference steps.

Image Generation: Lora generated more realistic images with good resolution in less inference steps. The performance is very good.

IMAGE MANIPULATION:

Image inpainting: Image inpainting is a technique used to reconstruct lost or damaged parts of an image. It involves filling in the missing areas with plausible content that matches the surrounding context. This can be done manually by an artist or automatically using computer algorithms.



<u>Control image net generation:</u> ControlNet is a Stable Diffusion-based neural network for image generation that allows users to control the final image generation through various techniques like pose, edge detection, depth maps, and more. Unlike conventional text-to-image or image-to-image models, ControlNet provides greater control over image generations, allowing users to customize the generation process and create images that meet their specific needs. ControlNet generates clean, anatomically accurate images based on human poses, even with poses where limbs are folded or not visible, ensuring faithful representation of the input poses.Users can easily condition the generation with different spatial contexts such as a depth map, a segmentation map, a scribble, keypoints, and so on, turning a cartoon drawing into a realistic photo with incredible coherence.

Left picture 21 savage right we have Narendra Modi which the image generated in the same pose as left picture.

Midjourney Ai:

Midjourney combines advanced machine learning algorithms with human creativity to produce stunning images based on textual descriptions. The images generated took a bit of more time to generated with the same inference steps lora and diffusion models are produced. The image quality is also not that good with the small inference steps.

prompt = "beautiful girls playing in the beach beside the beautiful white castle in the sunset "

Conclusion: These are the experiments I have done by using different image generation models. LoRA performed well comparatively remaining models also produced better images with prompt by taking more inference steps. Dreamshaper performed well with low inference steps with the production of good images. By using LoRA we have done image inpainting and image manipulation.