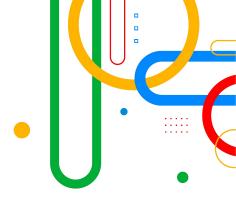


Google Cloud

Community Day 2023







Exploring GenAl using Diffusion Models

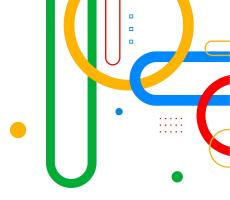






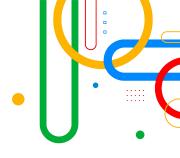
🖰 20th May, 2023 💿 Bengaluru











Community Day 2023

🖰 20th May, 2023 💿 Bengaluru



Speaker

Bismillah Kani

Designation

Staff Al/ML Scientist, Baker Hughes



Agenda

- Generative Al
- Generative Image Models
- Diffusion model
- Stable Diffusion model
- Prompt Engineering
- Demo App
- GenAl on GCP
- Summary





Generative Al

- Generative AI refers to a type of Artificial Intelligence that has the capability to generate new content and concepts, such as stories, conversations, videos, images and music.
- This technology relies on machine learning models, specifically Foundation Models (FMs), which are extensively trained on enormous dataset.



Text Generation



Image Generation

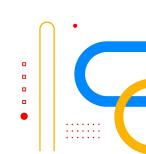


Code Generation



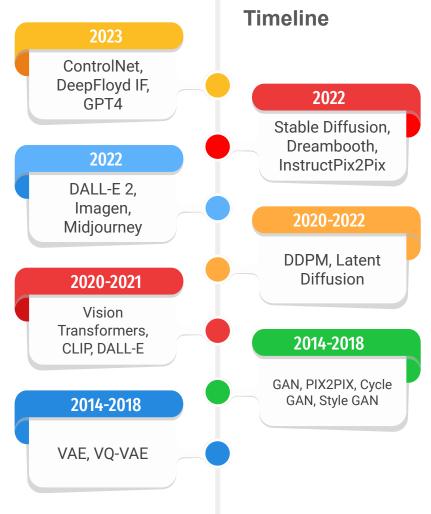
Virtual Assistant





Generative Image Models

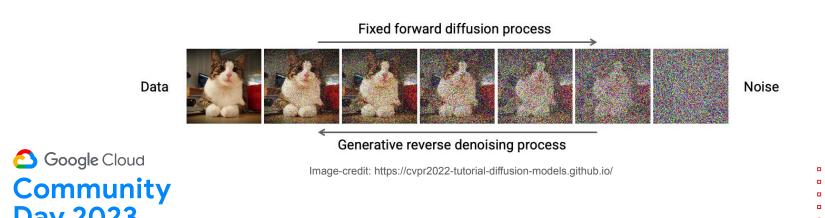
Generative image models utilize methods such as diffusion to create novel images. These models can generate new images by transforming random noise or by producing images based on given prompts or related visual information.





Diffusion models

- Diffusion models are iterative denoising autoencoders that progressively enhance an image to achieve a final, clean, and denoised output.
- This process starts with random noise and undergoes multiple steps of refinement.
- During each step, the model determines the optimal transformation from the current input to a denoised version.



Denoising Diffusion Probabilistic Model (DDPM)

In DDPM (Denoising Diffusion Probabilistic Model), the focus is on estimating the amount of noise that was introduced and iteratively removing that noise during inference. The model aims to progressively reduce the noise level and refine the image to its original form by iteratively applying noise removal operations.

The key components of this setup are as follows:

- 1. Scheduler: Introducing noise to the images during training. It controls the amount and type of noise added at each iteration or step of the training process.
- 2. Model: The model used in this setup is a modified version of the U-Net architecture.
- 3. Training: During the training phase, the model is trained to learn the relationship between the noisy input images and their corresponding clean counterparts.
- 4. Inference: In the inference phase, given a noisy image as input, the model applies its learned denoising capabilities to iteratively remove the noise and generate a cleaner version of the image as output.





Stable Diffusion Model

- Diffusion models can face challenges with generating high-resolution images due to increased computational requirements when processing larger images with U-Net architectures.
- A solution to this challenge involves performing diffusion operations in a latent space, utilizing an encoder-decoder framework for image conversion.
- By incorporating text conditioning, diffusion models can generate desired images based on specific textual prompts, rather than random image generation.
- Stable Diffusion, which utilizes these techniques, has achieved state-of-the-art results and can be
 deployed on consumer GPUs to produce high-quality images. The model was trained on a curated
 dataset of aesthetically pleasing images, specifically a subset of LAION 5B referred to as LAION
 aesthetics.





Stable Diffusion Model

Stable Diffusion Model operates as a system consisting of multiple interconnected models collaborating to generate high-quality images with conditioning.

The process involves the following components:

- Perceptual Image Compression: VAEs (Variational Autoencoders) are trained to efficiently compress and decompress images, enabling effective utilization of the latent space.
- Latent Diffusion: Instead of applying diffusion directly on the pixel level, diffusion operations are performed on the latent space. This helps reduce computational complexity while maintaining image quality.
- Conditioning: The CLIP (Contrastive Language-Image Pretraining) text encoder is employed to encode the text prompt. The resulting embedding is then concatenated and used for cross attention to condition the latent diffusion process.

Image-credit: https://learnopencv.com/stable-diffusion-generative-ai/

Community

Day 2023



Stable Diffusion Model



Futuristic cyborg, sleek metal enhancements and glowing circuits, standing in high-tech laboratory, intense, detailed, high resolution, sci-fi portrait





Giant caterpillar riding a bicycle, surreal, absurd, whimsical, stylized, detailed, vivid, high resolution, centralized, overhead view, colorful, fun



Delectable pizza with melted cheese, juicy tomato sauce, and an array of toppings including pepperoni, mushrooms, and black olives, served hot and fresh, high resolution, stylized, focused on ingredients and melted cheese

Prompt Engineering

Prompt engineering is the process of structuring words that can be interpreted and understood by a text-to-image model. Think of it as the language you need to speak in order to tell an Al model what to draw.

Structure of a good prompt

Subject

Desired content or elements to be depicted in the image

Medium

The material or medium utilized to create the artwork

Style

The artistic style or aesthetic approach desired for the image

Artist

Referencing the style of a specific artist as a point of inspiration

Resolution

Represents the level of sharpness and detail present in the image

Color

Exerting control over the overall color palette of the image

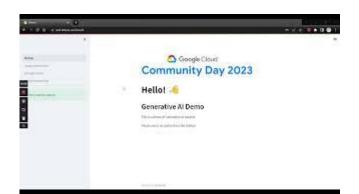
Lighting

Substantial impact on the visual appearance and ambiance of the image.



Community Day 2023

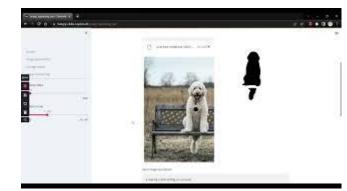
Demo App



Demo-1: Stable Diffusion



Demo-2: Al assisted Prompts



Demo-3: SAM assisted In-painting

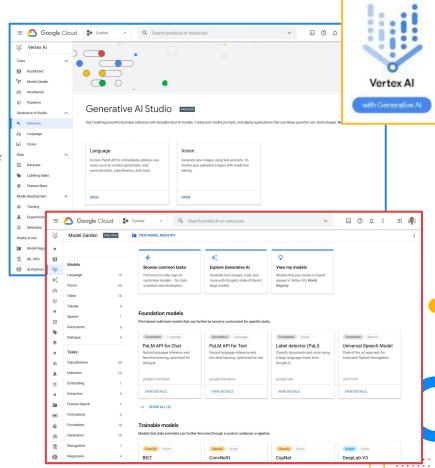


Community
Day 2023

GenAl on GCP

- In Vertex AI, the Generative AI Studio offers a managed environment designed to facilitate seamless interaction, tuning, and deployment of foundation models to production.
- In Vertex AI, the Model Garden serves as a comprehensive platform to accelerate your machine learning projects. It provides a centralized hub where you can search, explore, and engage with a diverse range of models developed by Google and its partners.





Summary



- Generative AI encompasses the field of creating new content, such as images, using artificial intelligence techniques.
- Generative Image Models have evolved over time, with diffusion models representing a departure from previous approaches like VAE and GAN, offering unique advantages.
- Latent Diffusion Model (LDM) and Stable Diffusion are powerful frameworks that employ diffusion techniques for generating high-quality images, including text-conditioned generation.
- Prompt engineering involves structuring language instructions to effectively communicate with text-to-image models, guiding them in generating desired content.
- Google Cloud's Generative Al Studio and Model Garden within Vertex Al provide managed environments for effortless interaction, tuning, and deployment of foundation models, accelerating GenAl projects.









Thank You

