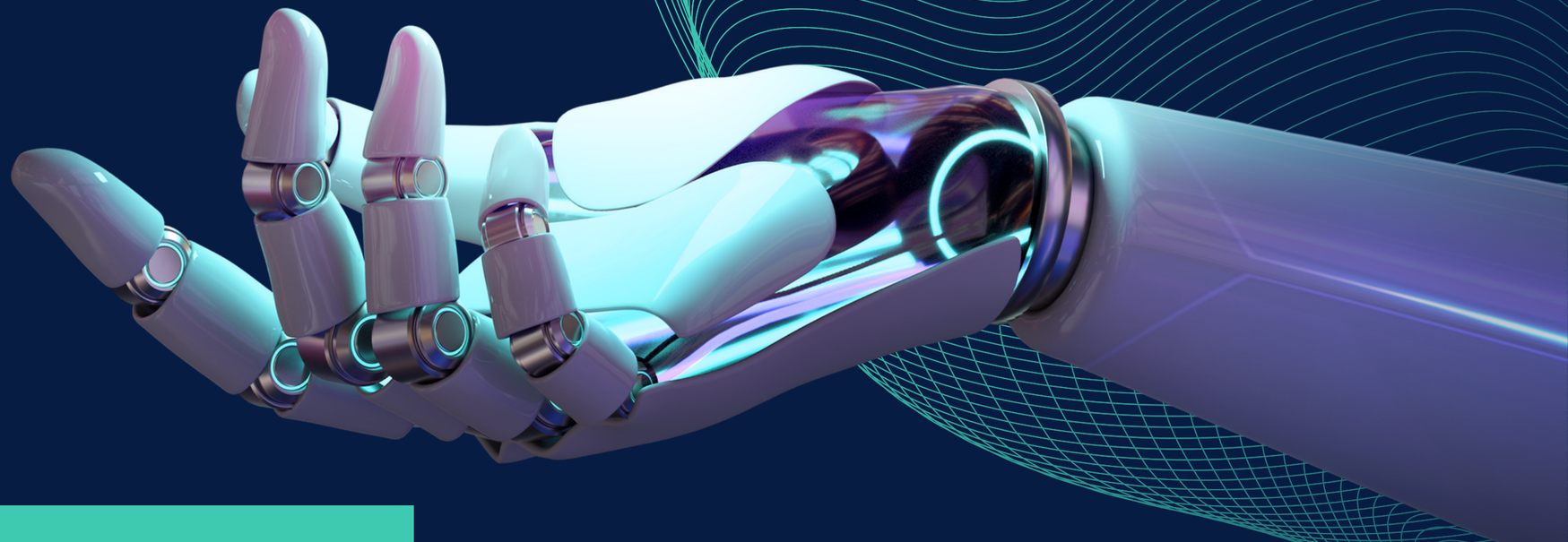
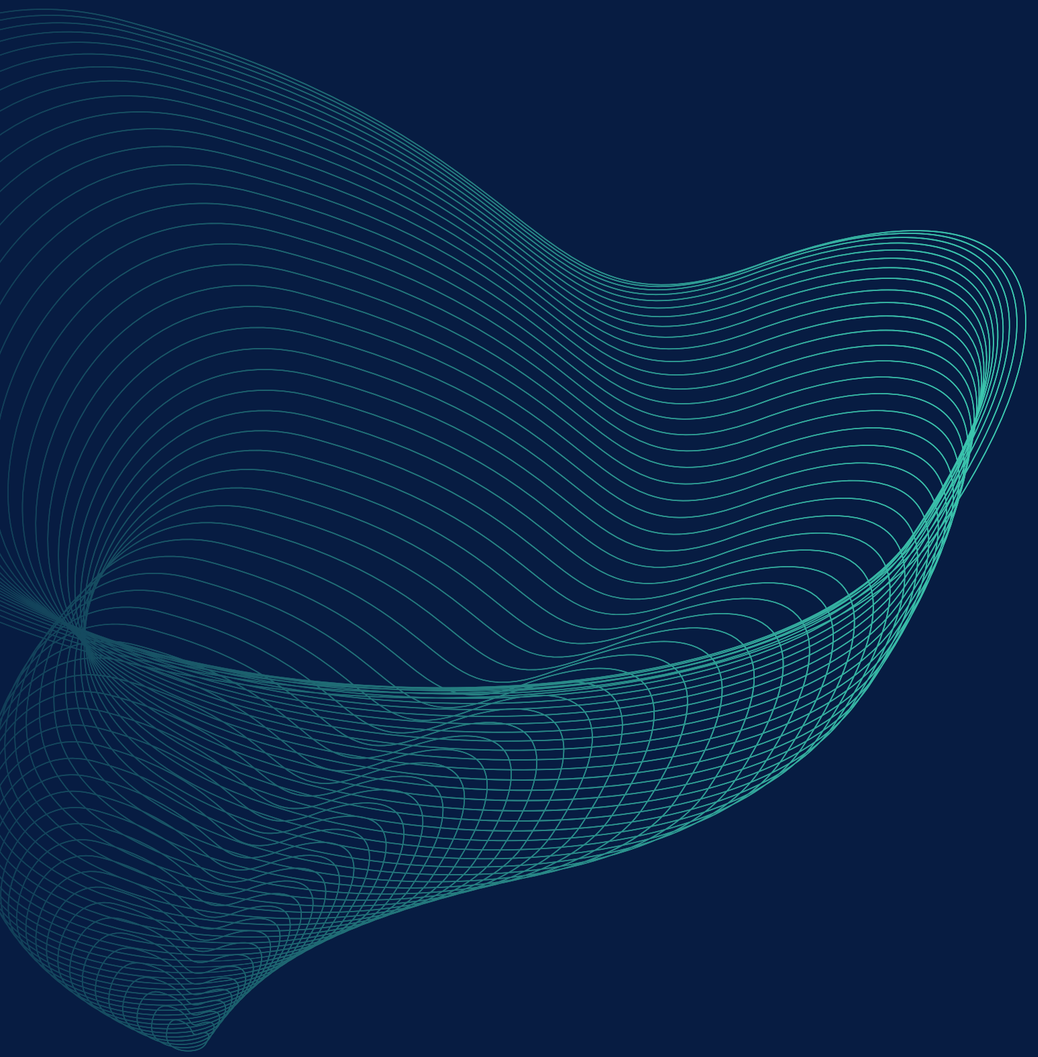
Exploring GenAI with Diffusion Model





Agenda

- Generative AI
- Diffusion Models
- Prompt Engineering
- Building GenAI

About me

I'm Bismillah Kani

Staff AI/ML Scientist at Waygate Technologies

AWS Community Builder - Machine Learning

AWS Certified SAA and MLS

/in/bismillah-kani

bismillahkani@gmail.com

https://github.com/bismillahkani

Generative AI

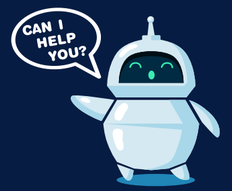
Generative AI refers to a type of artificial intelligence that has the capability to generate new content and concepts, such as stories, conversations, videos, images, and music.

This technology relies on machine learning models, specifically Foundation Models (FMs), which are extensively trained on enormous datasets.

Text Generation

Image Generation

Code Generation



Virtual Assistant

a brown leather jacket

Generative Image Models

These advancements pave the way for an exciting future in the field of generative AI, promising further innovations and breakthroughs.

2014-2018

VAE, VQ-VAE

2014-2018

GAN, PIX2PIX, Cycle GAN, Style GAN

2020-2021

Vision Transformers, CLIP, DALL-E

2020-2022

DDPM, Latent Diffusion

2022

DALL-E 2, Imagen, Midjourney

2022

Stable Diffusion, Dreambooth, InstructPix2Pix

2023

ControlNet, DeepFloyd IF, GPT4

Generative Image Models

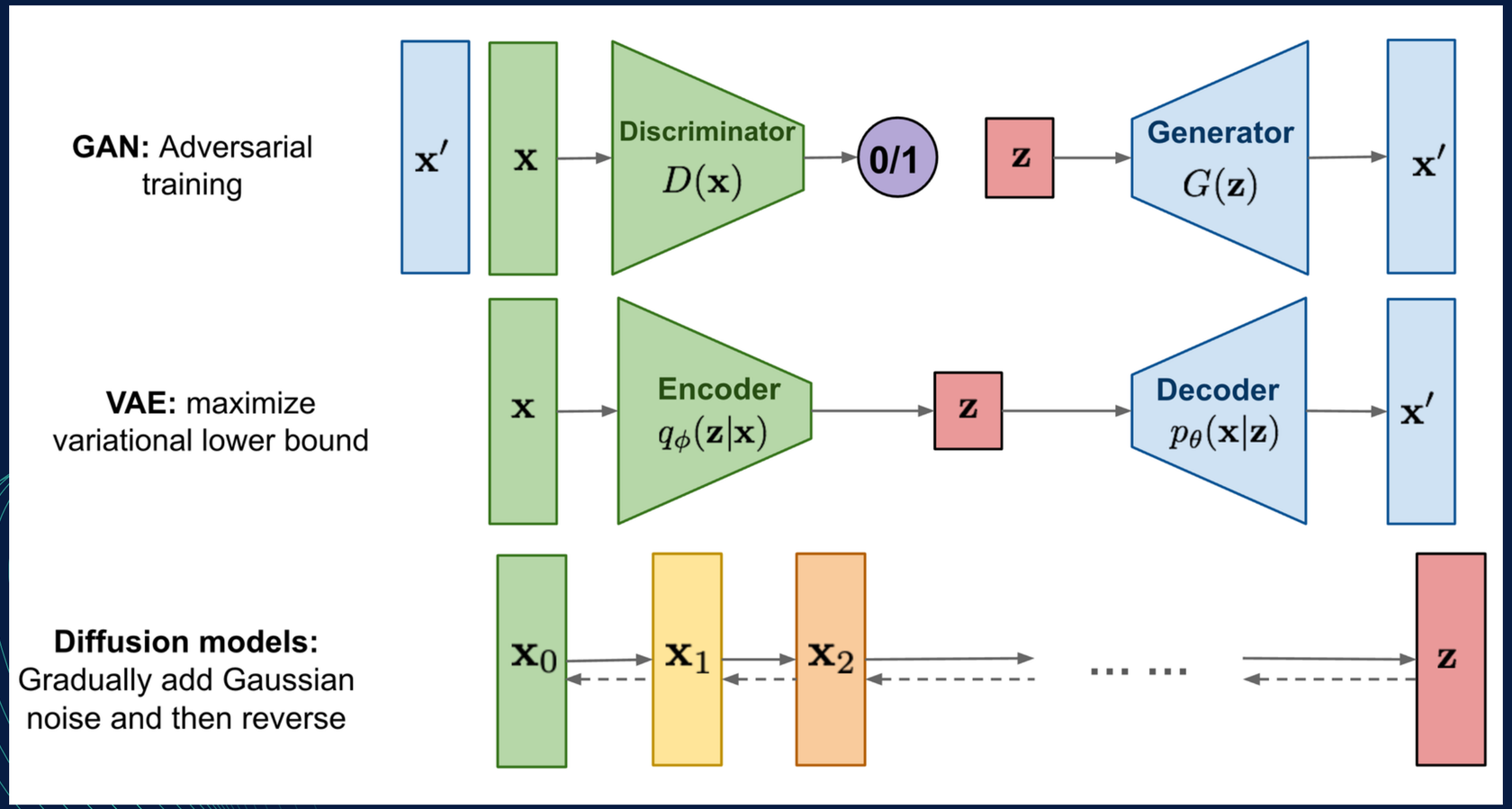


Image-credit: https://lilianweng.github.io/

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.

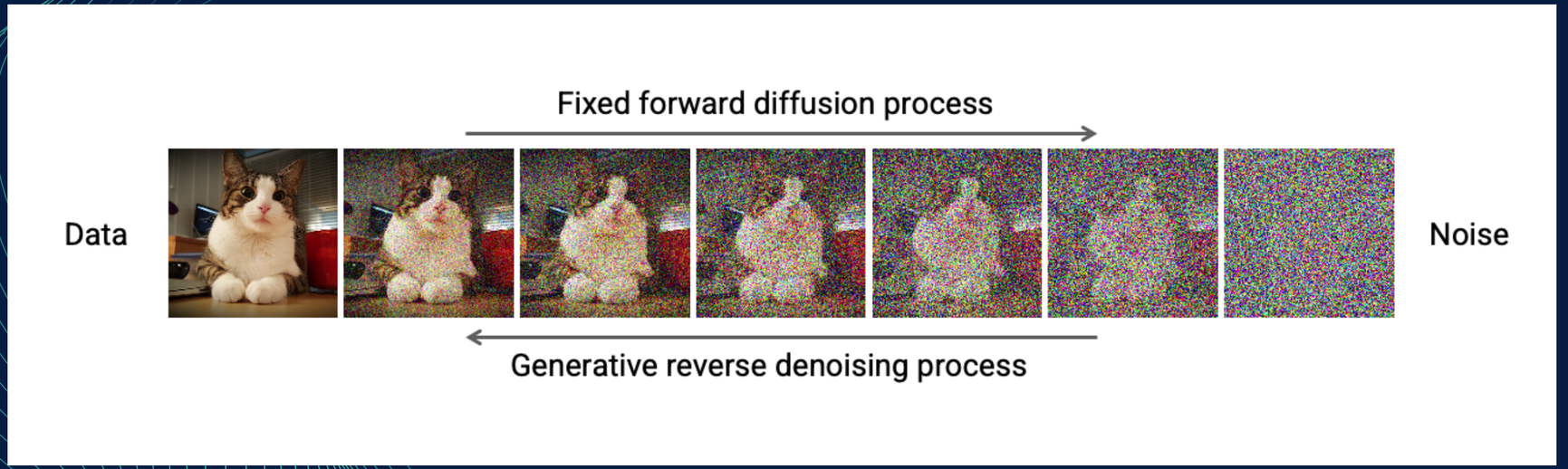


Image-credit: https://cvpr2022-tutorial-diffusion-models.github.io/

Stable Diffusion

- Diffusion models can face challenges with generating high-resolution images due to increased computational requirements when processing larger images with U-Net architectures.

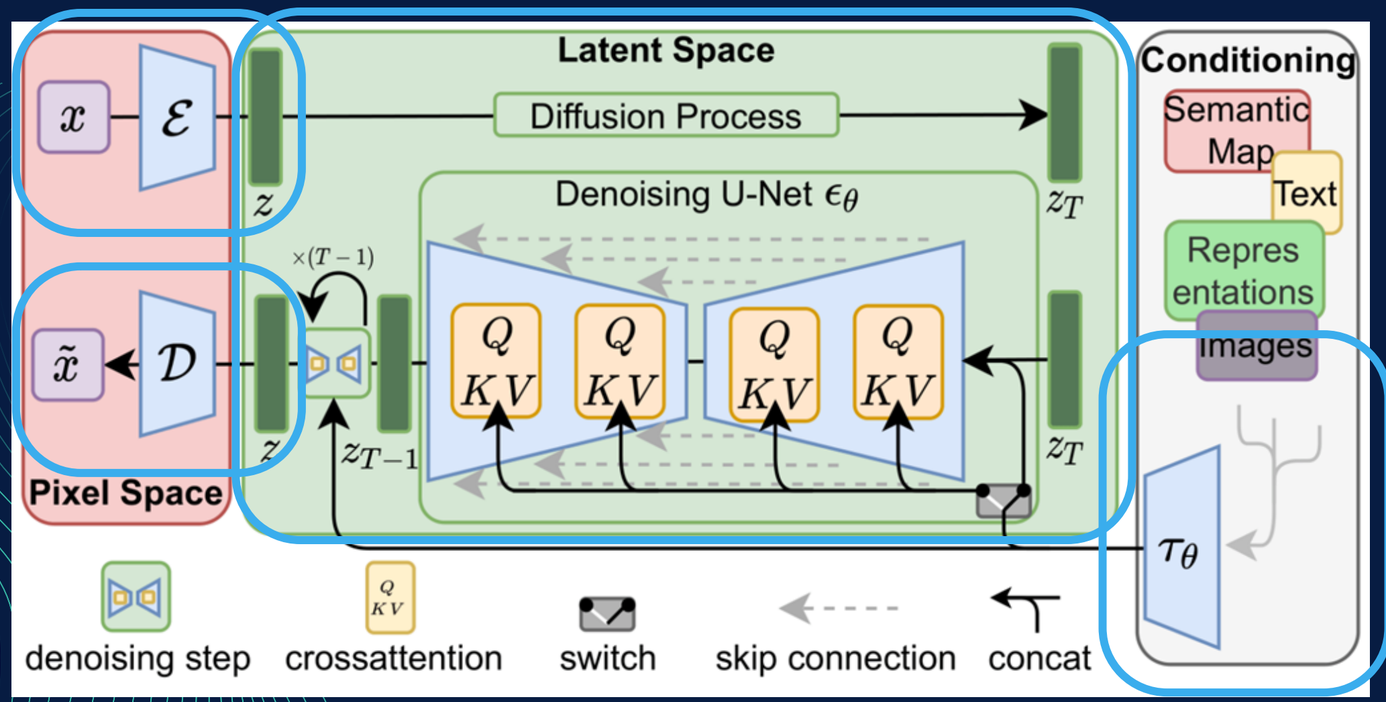
- A solution to this challenge involves performing diffusion operations in a latent space, utilizing an encoder-decoder framework for image conversion.

- 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

1. Encoder compress the input image into a 2D latent vector Z
2. Apply difusion and de-noising process on latent vector Z
3. Add conditioning via text encoder and cross-attention
4. Decoder reconstruct images from latent vector Z



Prompt Engineering

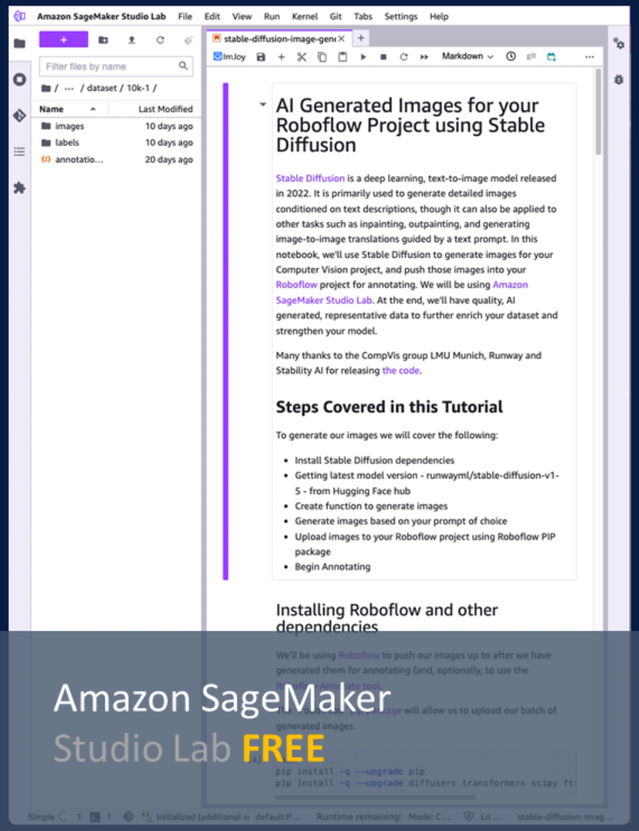
Prompt engineering is the process of structuring words that can be interpreted and understood by a text-to-image model. Think of it as the language you need to speak in order to tell an AI model what to draw.

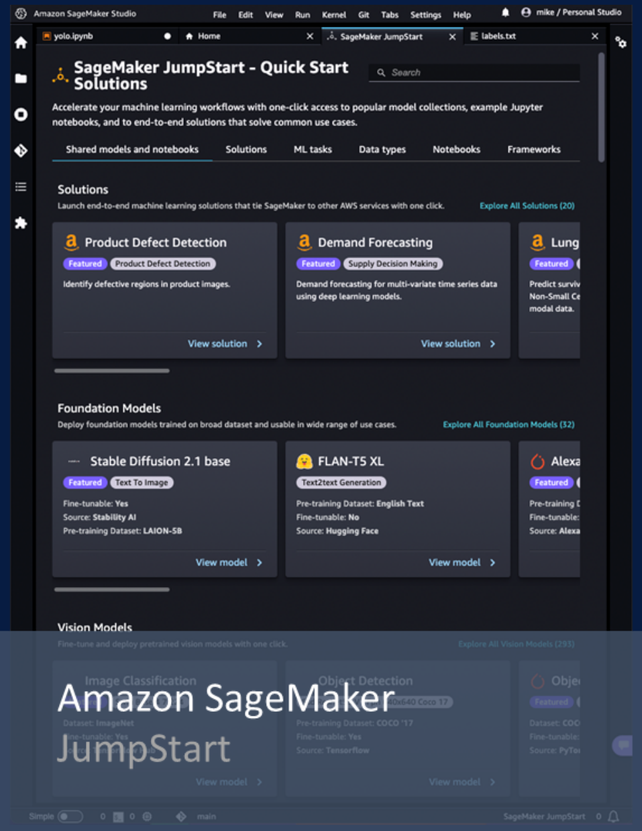
SUBJECT	MEDIUM	STYLE	ARTIST	RESOLN	COLOR	LIGHTNING
Desired content or elements to be depicted in the image	The material or medium utilized to create the artwork	The artistic style or aesthetic approach desired for the image	Referencing the style of a specific artist as a point of inspiration	Represents the level of sharpness and detail present in the image	Exerting control over the overall color palette of the image	Substantial impact on the visual appearance and ambiance of the image.

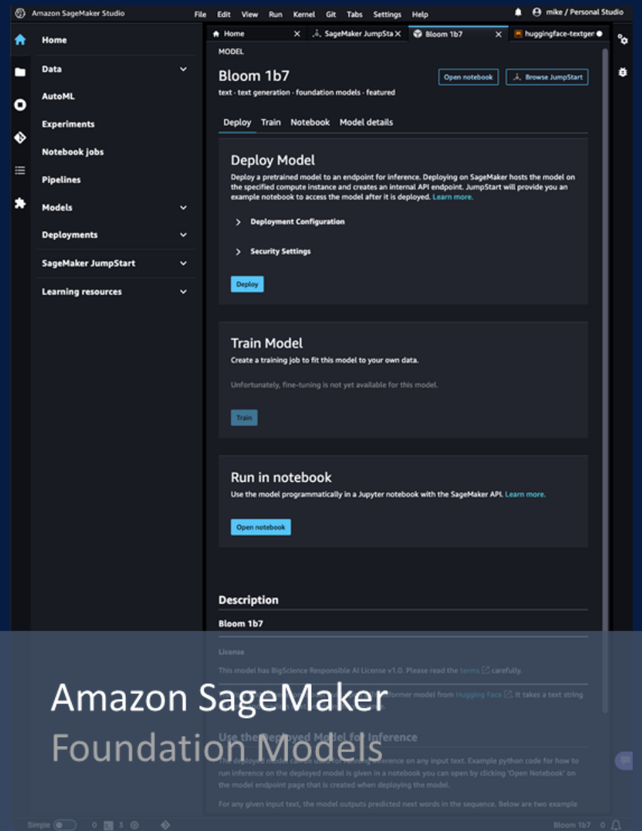
SageMaker Jumpstart

JumpStart is the machine learning (ML) hub of SageMaker that provides hundreds of built-in algorithms, pre-trained models, and end-to-end solution templates to help you quickly get started with ML.

To utilize a big model like Stable Diffusion on Amazon SageMaker, JumpStart provides a simplfied process, by offering pre-tested, readily available scripts accessible through the Studio UI with a single click or through the JumpStart APIs with minimal code.







SageMaker Jumpstart

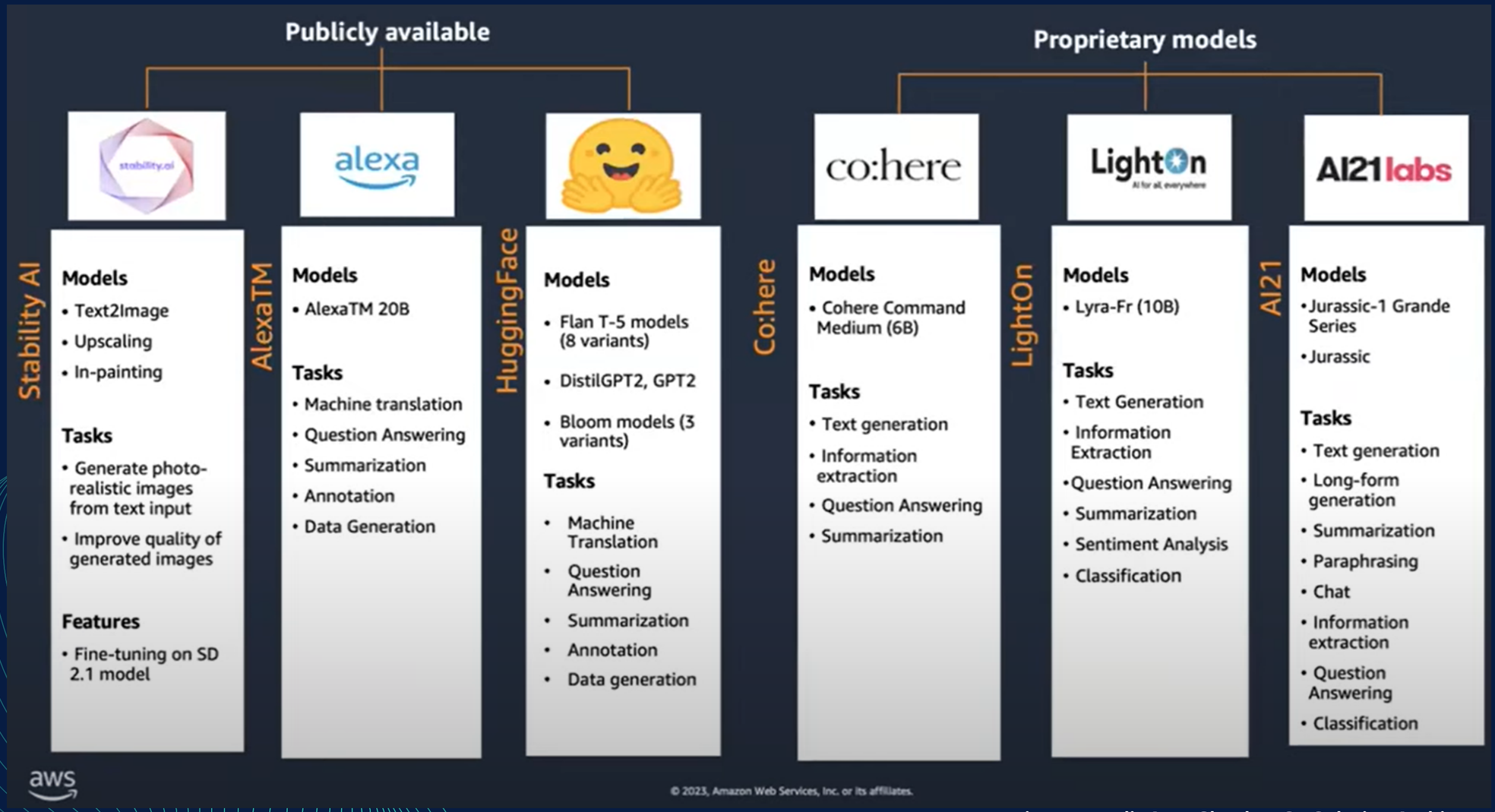
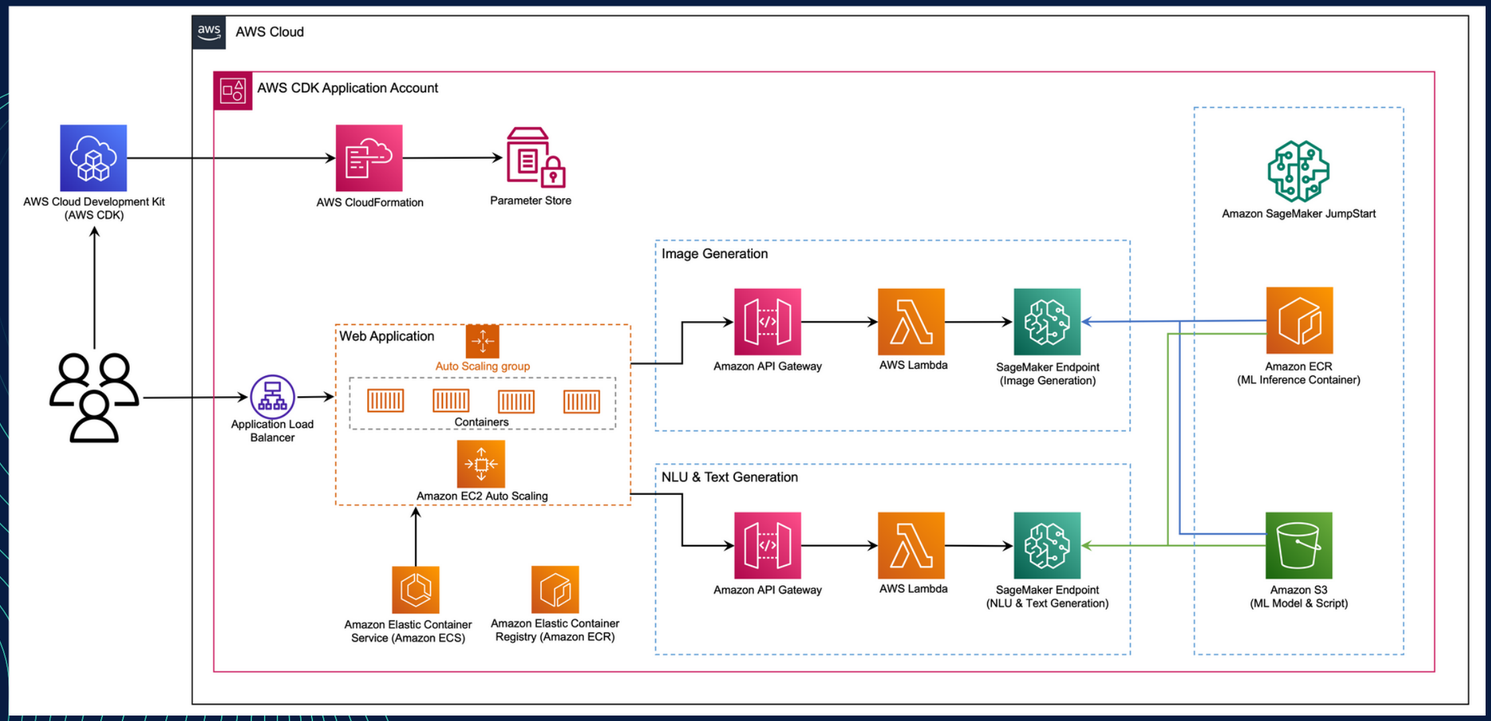


image-credit:Arun Shankar, Sr. Solution Architect @ AWS AI/ML

Demo App

- The web application is created using Streamlit, a Python library that facilitates the development and sharing of customized web apps for machine learning and data science.
- To host the web application, we utilize Amazon Elastic Container Service (Amazon ECS) in conjunction with AWS Fargate, which allows for container execution without the need to manage servers, clusters, or virtual machines.
- The generative AI model endpoints are launched via SageMaker Jumpstart images stored in Amazon Elastic Container Registry (Amazon ECR).
- The interaction between the web application and models takes place through Amazon API Gateway and AWS Lambda functions, as depicted in the diagram below.



Thank you