Generative AI for Image Generation

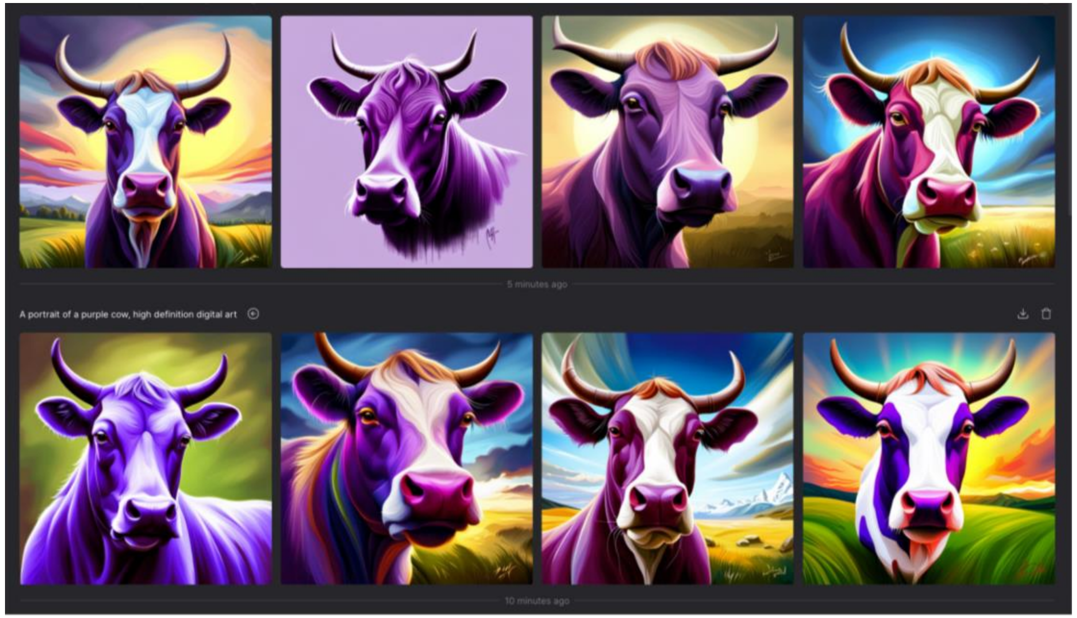
Day 2

Outlines

- Generative AI Platforms for Image Generation
- How Stable Diffusion model works?
- Localized Stable Diffusion App
- Image Generation using APIs
- Integrated Prompt Generation with Image Generation with APIs
- Foundation Models

Image Generation

https://platform.stability.ai/



5 minutes ago

A portrait of a purple cow, high definition digital art

10 minutes ago

https://zapier.com/blog/how-to-use-stable-diffusion/

https://prog.world/sherudim-under-the-hood-of-stable-diffusion/

Video Generation

https://openai.com/index/sora/

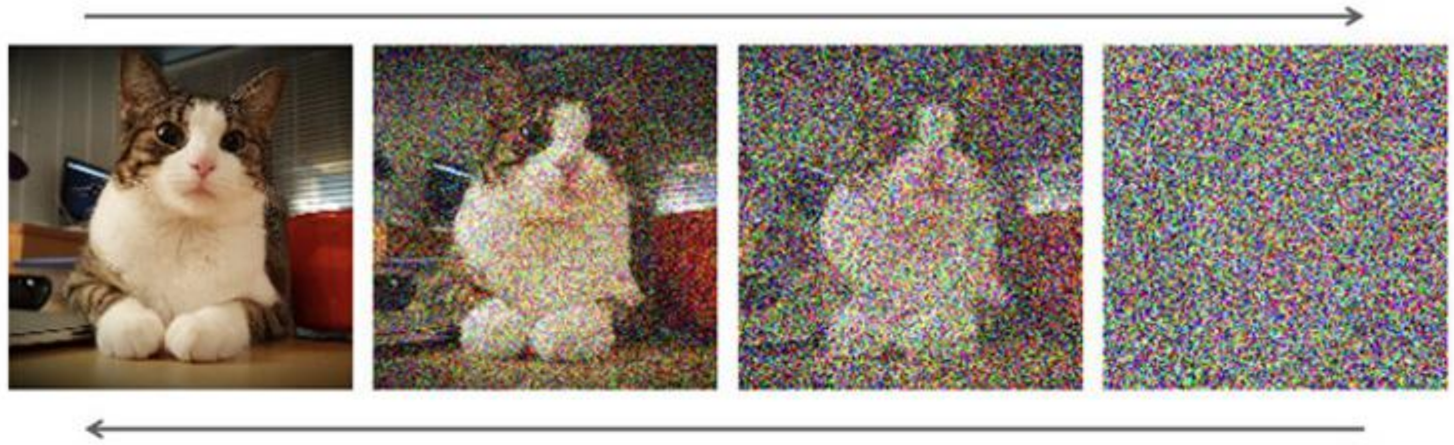
https://www.youtube.com/watch?v=HK6y8DAPN_0&t=43s

Generative AI Platforms for Image Generation

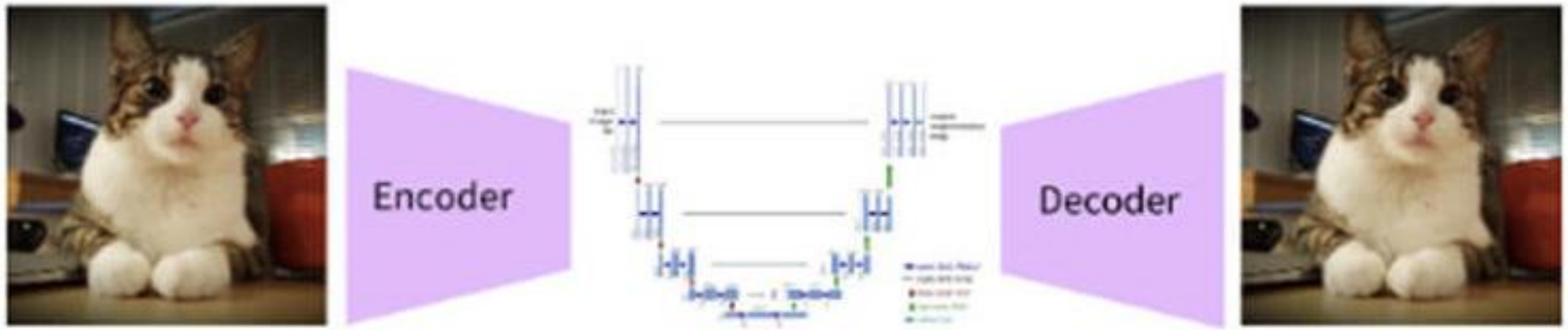
	Dall-E	Midjourney	Stable Diffusion
Website	https://openai.com/index/dall-e-3	https://www.midjourney.com/showcase	https://stability.ai/stable-image
Architecture	Transformer + Diffusion Model		
Ways to use	• Cloud service • API	• Cloud service through Discord • API	• Cloud service • API • Customized on **local hardware**
General comparison	Likely to make the most accurate semantic interpretation and interpolation judgments.	Produce the best-looking images even without sophisticated prompts.	Act more consistently and make fewer errors.
Unique features	• Produce photorealistic images in a wide variety of styles • AI model scores high on visual reasoning tests designed for humans • Can expand an existing image beyond its original borders in a consistent way	• Create very sharp and detailed images that look highly realistic • Produce great-looking results even with vaguely defined prompts	• Produce original and detailed work that meets the technical requirements • Can redraw existing images with contextual changes requested • Possible to directly improve colors, textures, and other visual elements
Observed problems	• Often fail to establish proper relations between multiple objects in the image • Poorly suited for handling scientific images that depend on the exactness	• Take a relatively long time • Sometimes ignore technical instructions to create a 'prettier' image	• Occasionally generate images that are identical to those from its training set • No strict controls preventing violent or sexual images from being generated
Pricing	https://openai.com/api/pricing/ DALL·E 3, 1024×1024, $0.040 / image	$10, $30, and $60 per month	$9, $49, and $149 per month

Stable Diffusion Model

Forward diffusion process is the process where more and more noise is added to the picture. Therefore, the image is taken and the noise is added in t different temporal steps where in the point T, the whole image is just the noise.

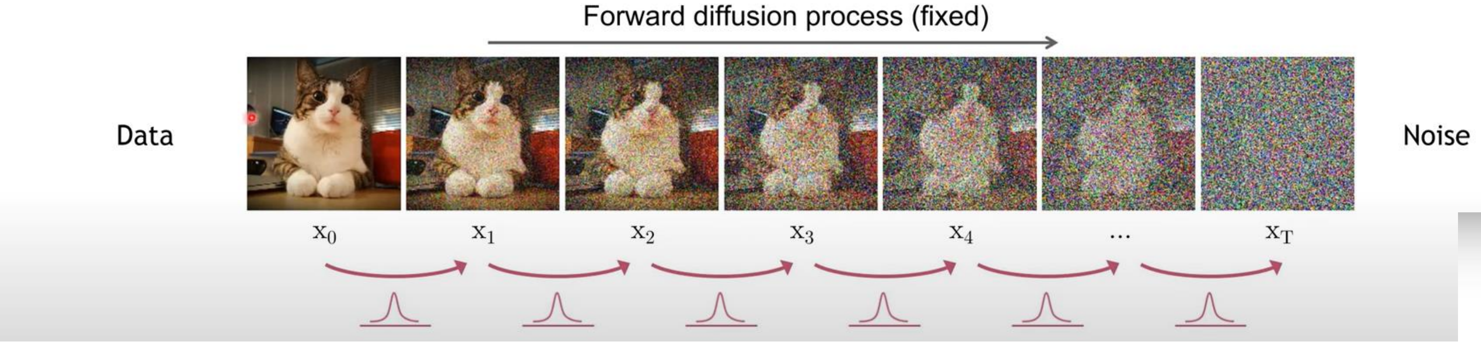


Backward diffusion is a reversed process when compared to forward diffusion process where the noise from the temporal step t is iteratively removed in temporal step t-1. This process is repeated until the entire noise has been removed from the image.

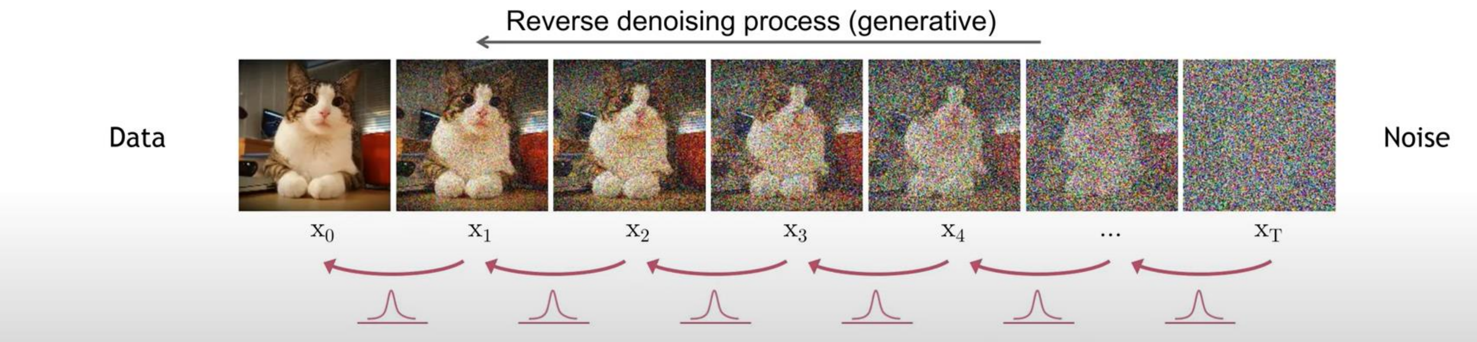


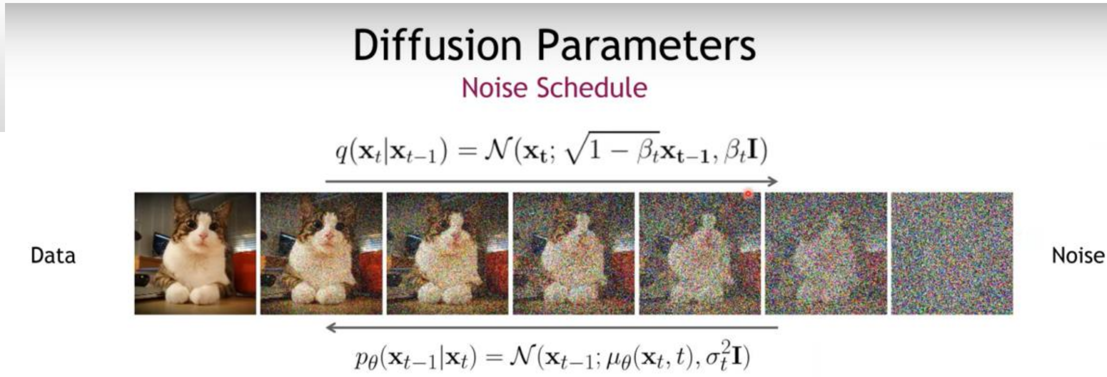
The architecture of the stable diffusion model

The formal definition of the forward process in T steps:



Formal definition of forward and reverse processes in T steps:





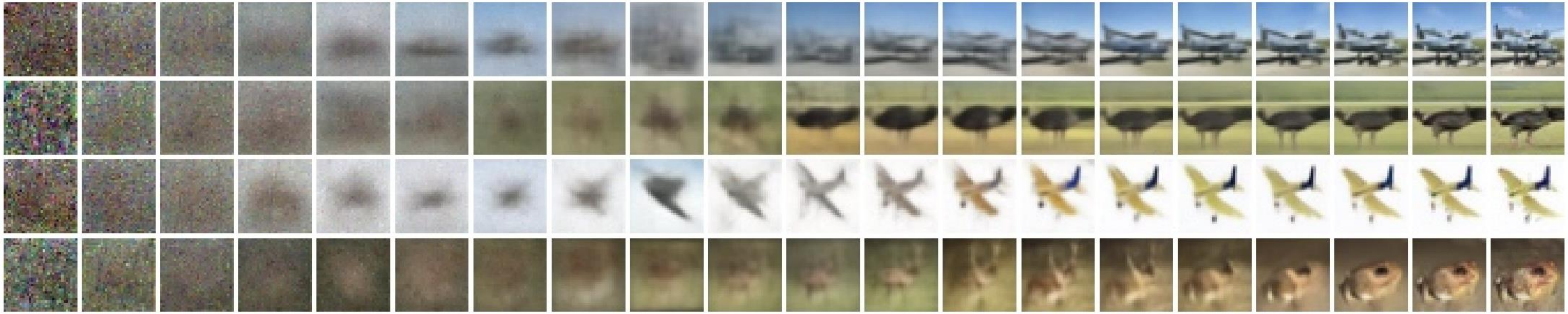
Diffusion Parameters

Noise Schedule

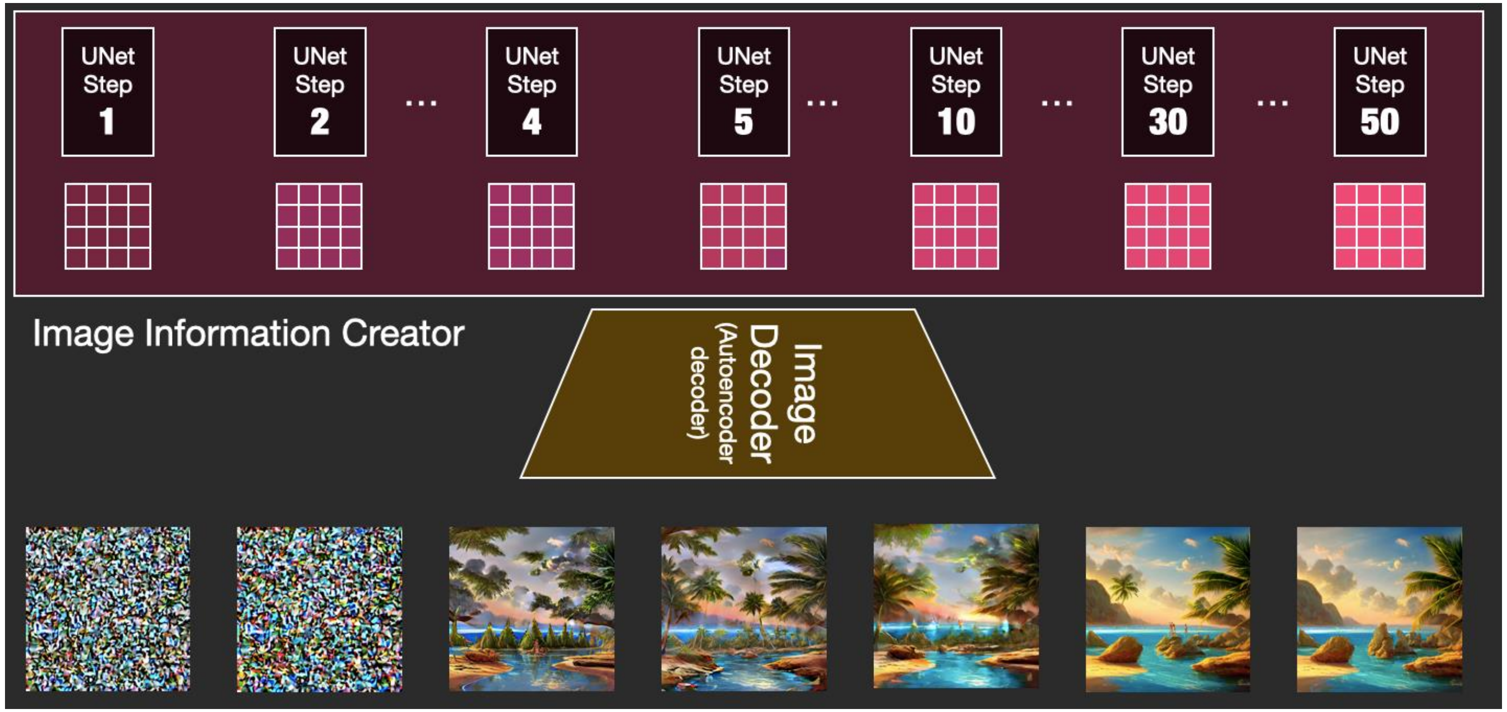
$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x_t}; \sqrt{1-\beta_t}\mathbf{x_{t-1}}, \beta_t\mathbf{I})$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2\mathbf{I})$$

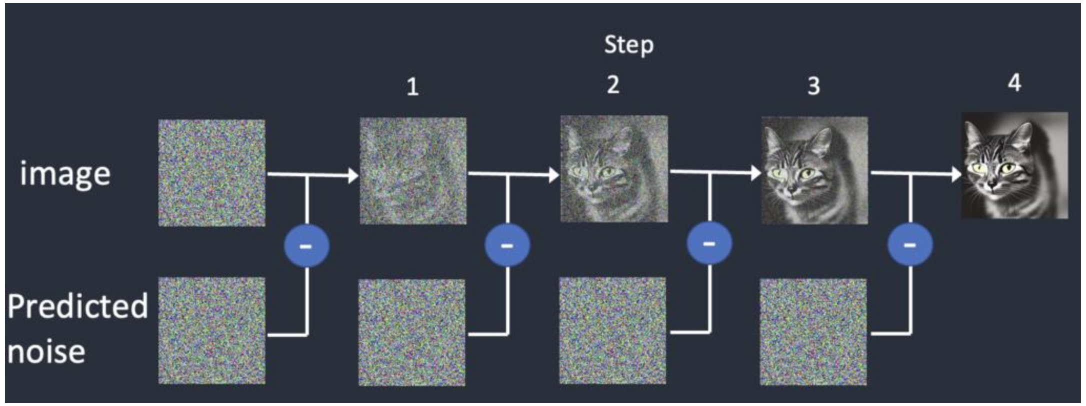
Parametrizing the Denoising Model

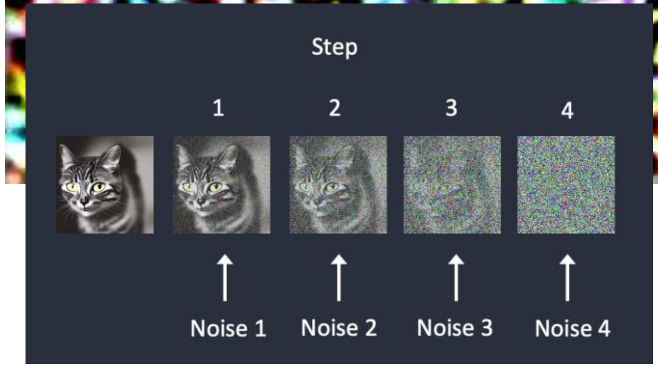


Train Model using large datasets

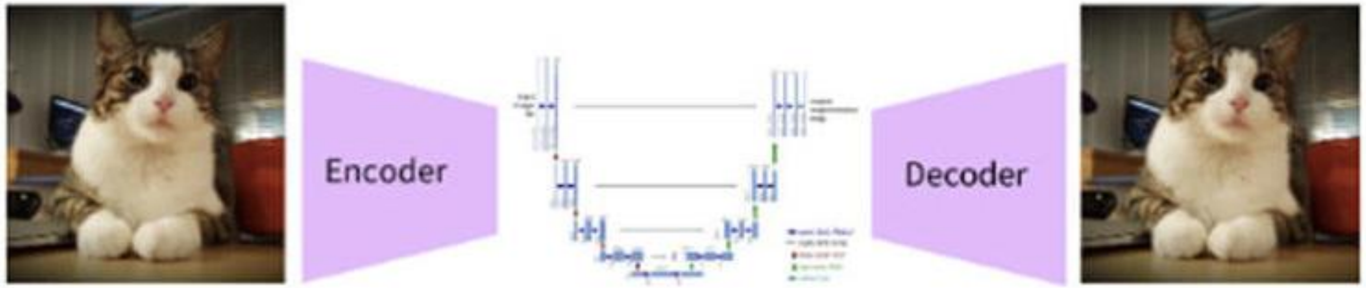


The model generates images by iteratively denoising random noise until a configured number of steps have been reached, guided by the CLIP text encoder pretrained on concepts along with the attention mechanism, resulting in the desired image depicting a representation of the trained concept.





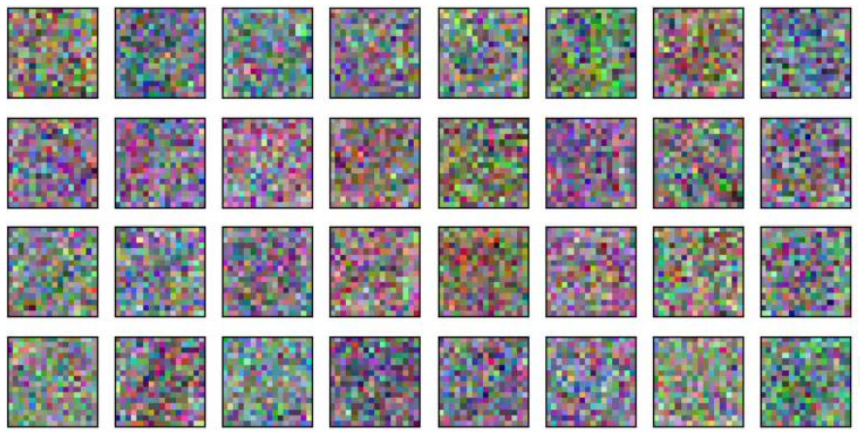
How Diffusion Model Works?

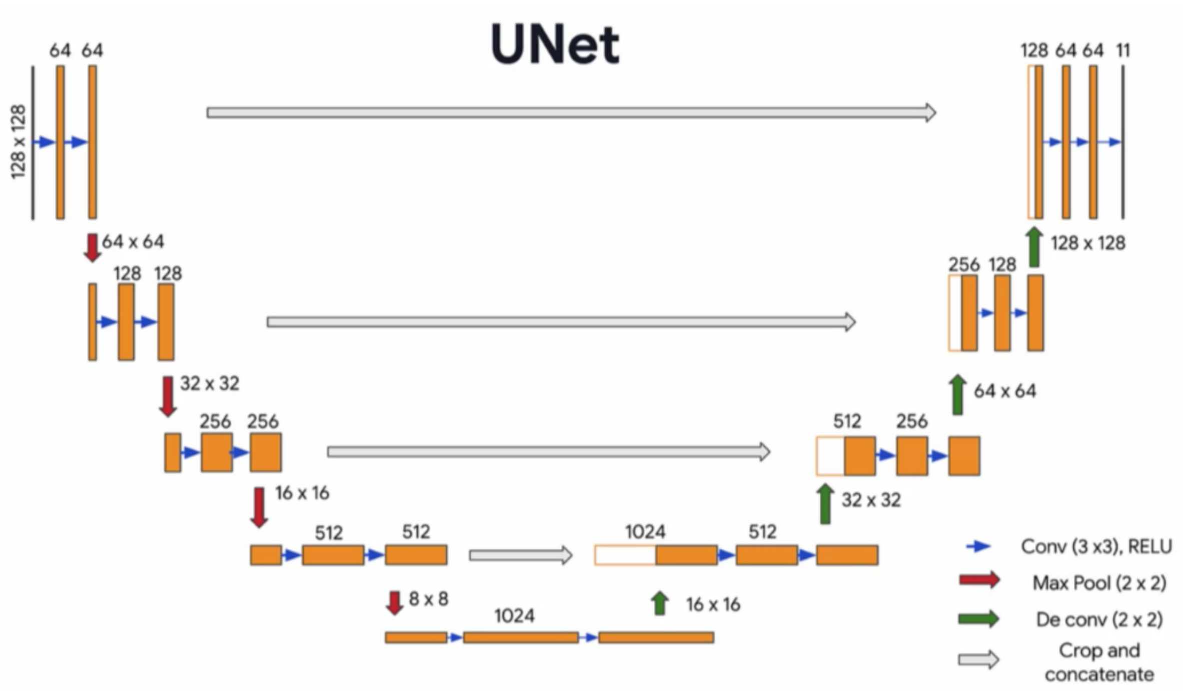


The architecture of the stable diffusion model

Demo in Notebook
How_Diffusion_Model_Work.ipynb

```
.n_channels, n_feat=256, n_cfeat=10, height=28):  # cfeat -
xtUnet, self).__init__()

 input channels, number of intermediate feature maps and

ınels = in_channels
 = n_feat
: = n_cfeat
        self.h = height  #assume h == w. must be divisible by 4, so
28,24,20,16...

itialize the initial convolutional layer
.init_conv = ResidualConvBlock(in_channels, n_feat, is_res=True)

itialize the down-sampling path of the U-Net with two levels
.down1 = UnetDown(n_feat, n_feat)         # down1 #[10, 256, 8, 8]
.down2 = UnetDown(n_feat, 2 * n_feat)     # down2 #[10, 256, 4,  4]

riginal: self
.to_vec = nn.

bed the times
ural network
.timeembed1 =
.timeembed2 =
.contextembed
.contextembed

itialize the
.up0 = nn.Seq
nn.ConvTransp

nn.GroupNorm(
nn.ReLU(),
        )
        self.up1 = UnetUp(4 * n_feat, n_feat)
```

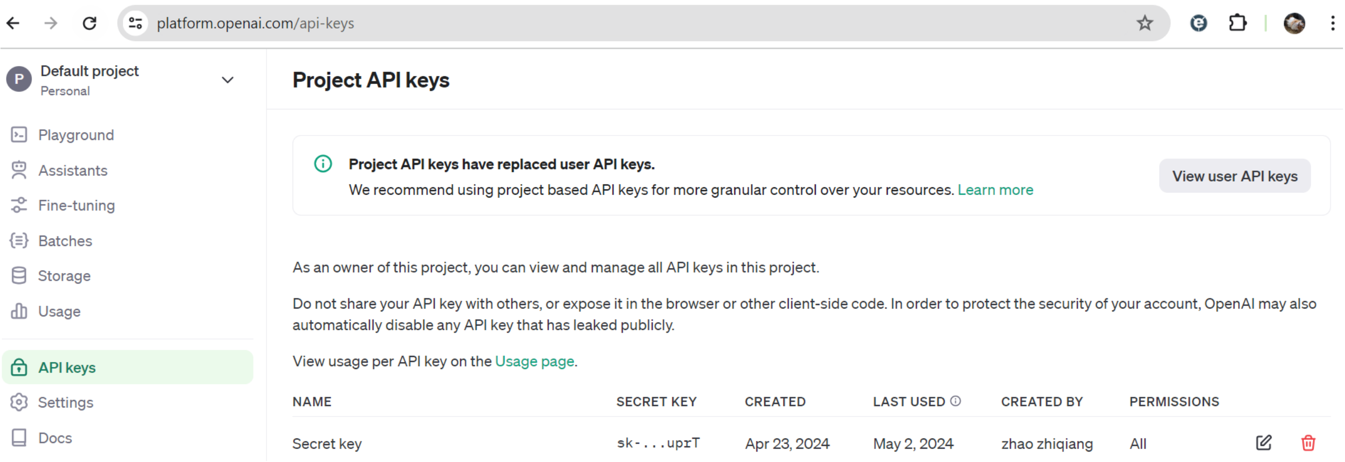




UNet Architecture

DALL·E

- https://openai.com/index/dall-e-3/
  - Require ChatGPT Plus
- Copilot (microsoft.com)
  - The free version of Copilot allows users to generate up to 30 images per day
- OpenAI API Keys
  - https://platform.openai.com/api-keys



Project API keys

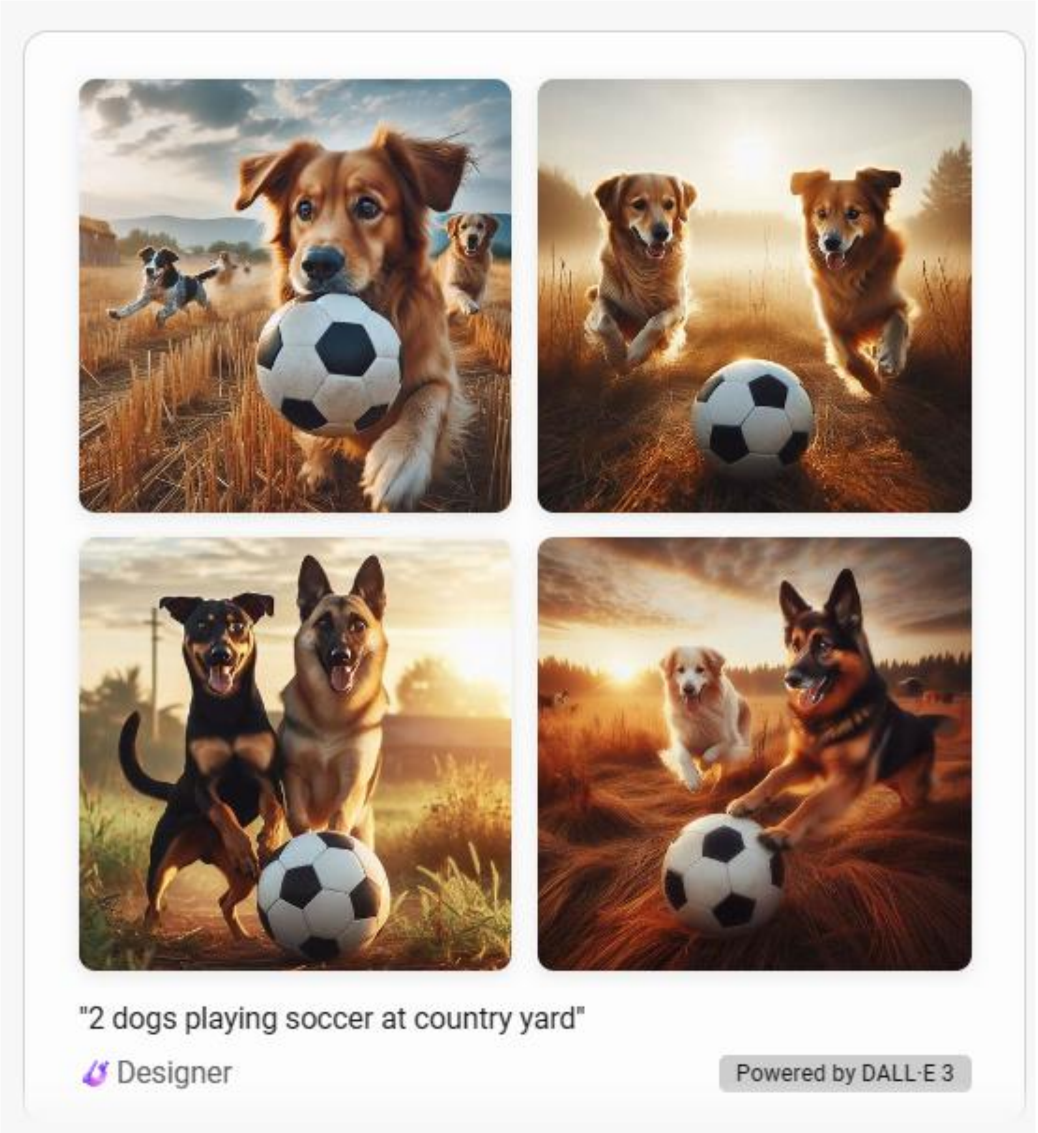
Project API keys have replaced user API keys.
We recommend using project based API keys for more granular control over your resources. Learn more

As an owner of this project, you can view and manage all API keys in this project.

Do not share your API key with others, or expose it in the browser or other client-side code. In order to protect the security of your account, OpenAI may also automatically disable any API key that has leaked publicly.

View usage per API key on the Usage page.

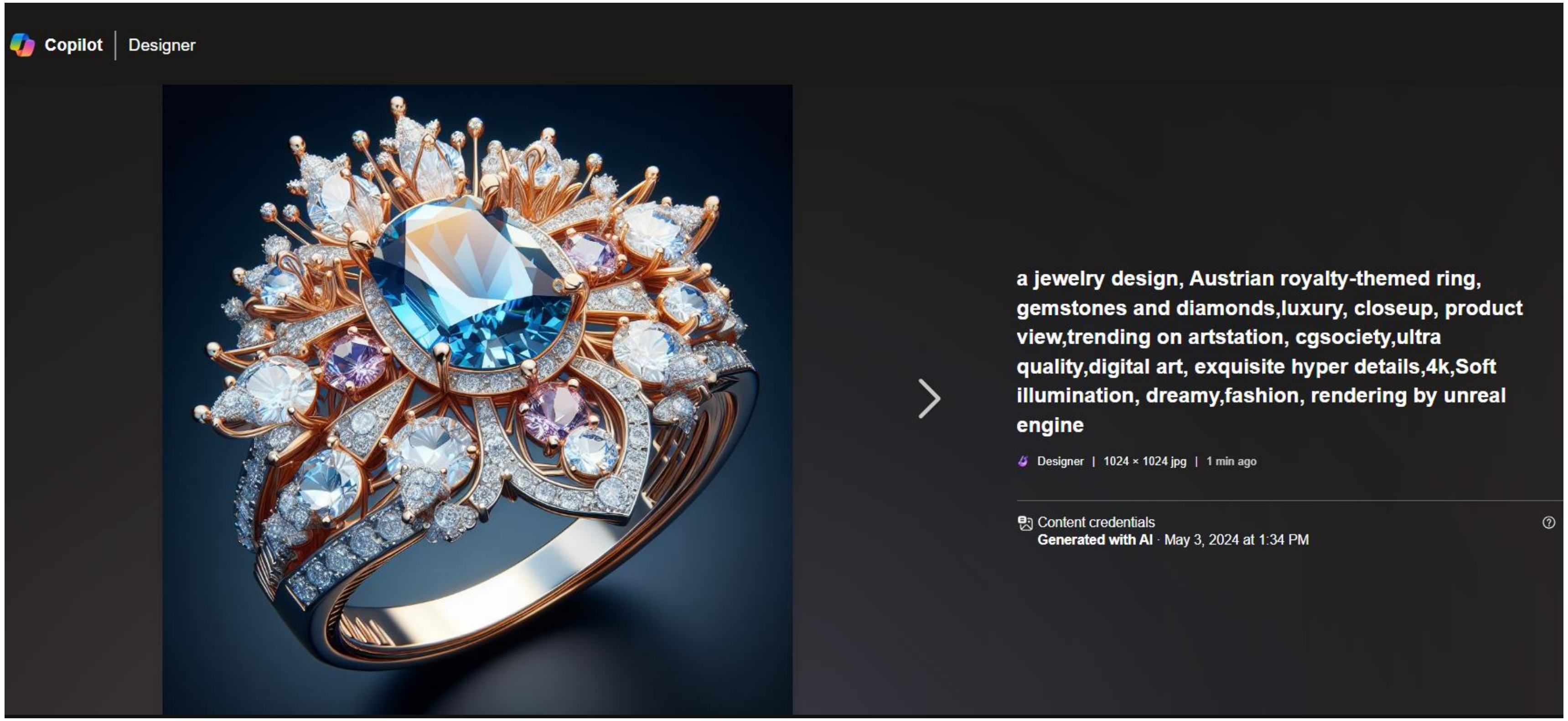
NAME	SECRET KEY	CREATED	LAST USED	CREATED BY	PERMISSIONS
Secret key	sk-...uprT	Apr 23, 2024	May 2, 2024	zhao zhiqiang	All


"2 dogs playing soccer at country yard"

Copilot – Generate Image

https://copilot.microsoft.com/

Prompt: “a jewelry design, austrian royalty-themed ring, gemstones and diamonds, luxury, closeup, product view, trending on artstation, cgsociety, ultra quality, digital art, exquisite hyper details,4k,Soft illumination, dreamy, fashion, rendering by unreal engine.”



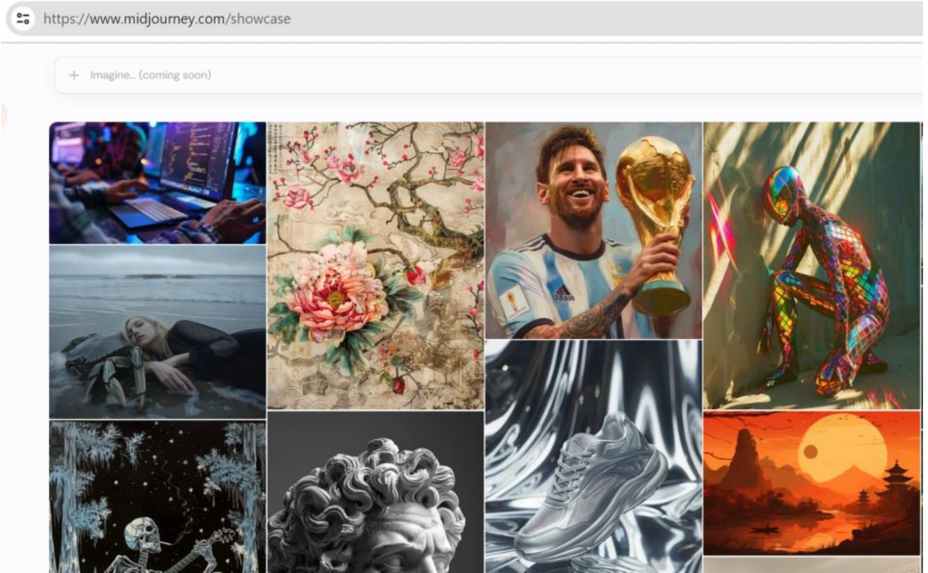
a jewelry design, Austrian royalty-themed ring, gemstones and diamonds,luxury, closeup, product view,trending on artstation, cgsociety,ultra quality,digital art, exquisite hyper details,4k,Soft illumination, dreamy,fashion, rendering by unreal engine

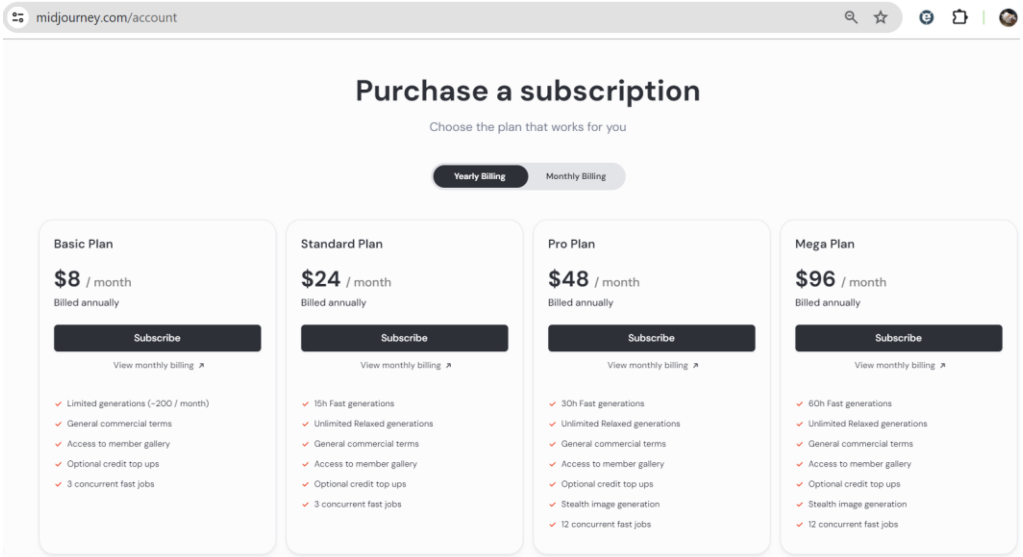
Designer | 1024 × 1024 jpg | 1 min ago

Content credentials
Generated with AI · May 3, 2024 at 1:34 PM

Midjourney

- Midjourney membership
  - https://www.midjourney.com/account
  - As of June 2023, Midjourney is no longer available for free.
- Generate images at https://discord.com/channels/
- API Keys
  - Purchase a subscription for Discord account on the midjourney.com website.


https://www.midjourney.com/showcase
+ Imagine... (coming soon)



Purchase a subscription

Choose the plan that works for you

Basic Plan

$8 / month

Billed annually

- ✓ Limited generations (~200 / month)
- ✓ General commercial terms
- ✓ Access to member gallery
- ✓ Optional credit top ups
- ✓ 3 concurrent fast jobs

Standard Plan

$24 / month

Billed annually

- ✓ 15h Fast generations
- ✓ Unlimited Relaxed generations
- ✓ General commercial terms
- ✓ Access to member gallery
- ✓ Optional credit top ups
- ✓ 3 concurrent fast jobs

Pro Plan

$48 / month

Billed annually

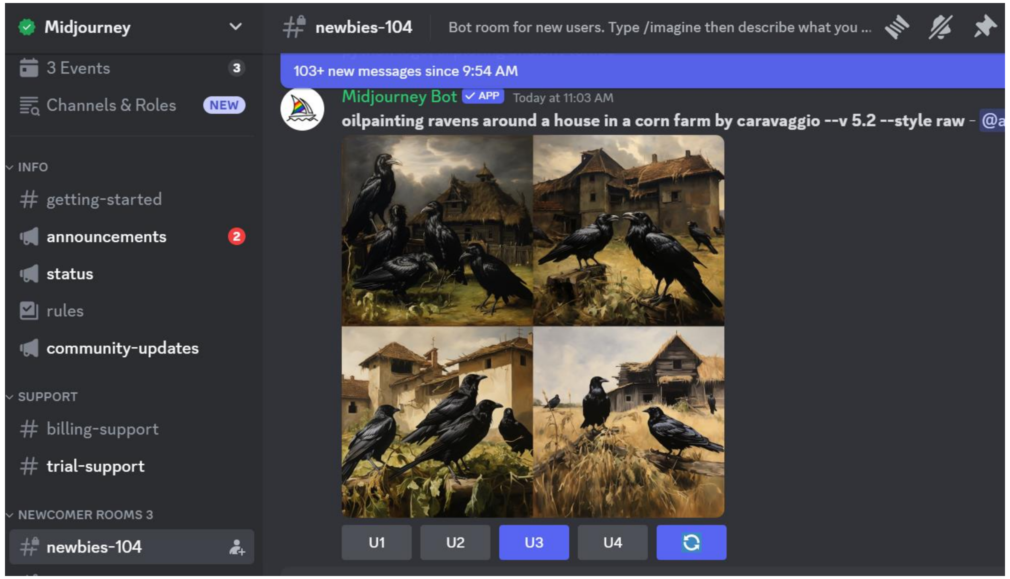
- ✓ 30h Fast generations
- ✓ Unlimited Relaxed generations
- ✓ General commercial terms
- ✓ Access to member gallery
- ✓ Optional credit top ups
- ✓ Stealth image generation
- ✓ 12 concurrent fast jobs

Mega Plan

$96 / month

Billed annually

- ✓ 60h Fast generations
- ✓ Unlimited Relaxed generations
- ✓ General commercial terms
- ✓ Access to member gallery
- ✓ Optional credit top ups
- ✓ Stealth image generation
- ✓ 12 concurrent fast jobs



# newbies-104 Bot room for new users. Type /imagine then describe what you ...

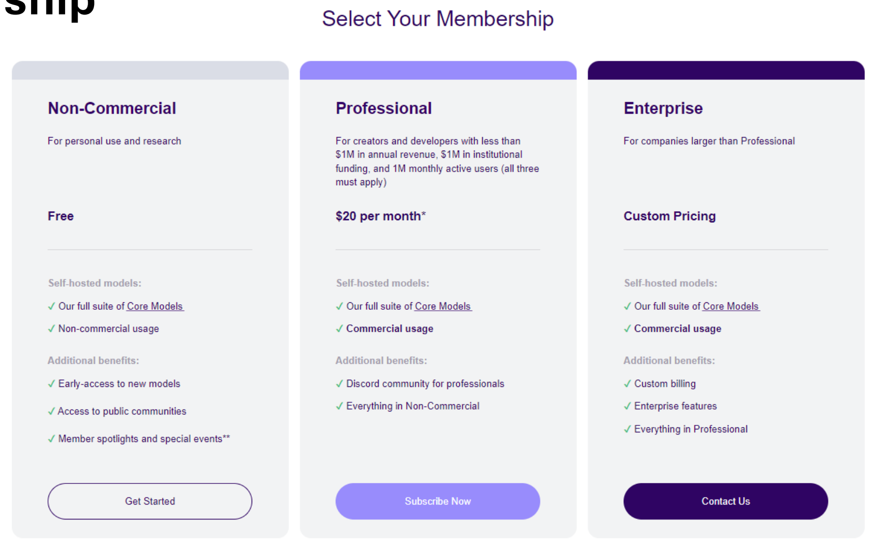
103+ new messages since 9:54 AM

Midjourney Bot ✓ APP Today at 11:03 AM

oilpainting ravens around a house in a corn farm by caravaggio --v 5.2 --style raw - @a

Stable Diffusion

Membership



Select Your Membership

Non-Commercial

For personal use and research

Free

Self-hosted models:

- ✓ Our full suite of Core Models
- ✓ Non-commercial usage

Additional benefits:

- ✓ Early-access to new models
- ✓ Access to public communities
- ✓ Member spotlights and special events**

Professional

For creators and developers with less than $1M in annual revenue, $1M in institutional funding, and 1M monthly active users (all three must apply)

$20 per month*

Self-hosted models:

- ✓ Our full suite of Core Models
- ✓ **Commercial usage**

Additional benefits:

- ✓ Discord community for professionals
- ✓ Everything in Non-Commercial

Enterprise

For companies larger than Professional

Custom Pricing

Self-hosted models:

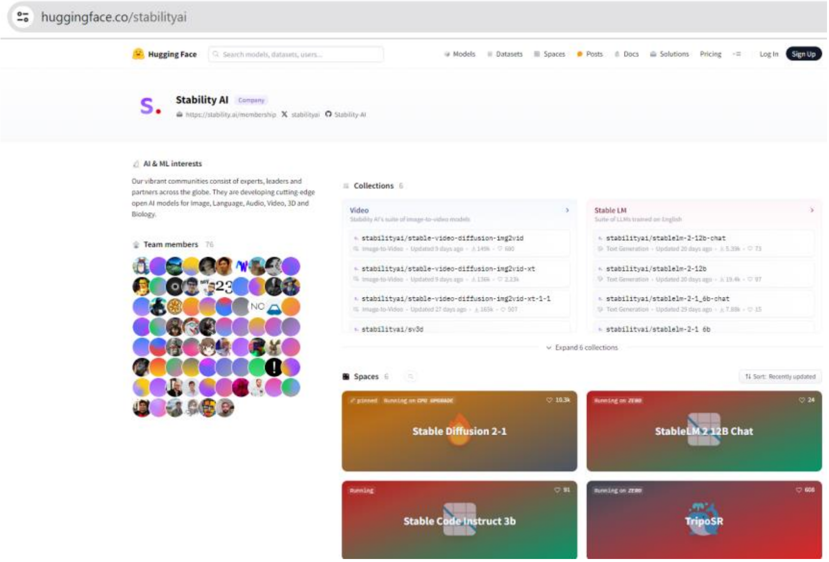
- ✓ Our full suite of Core Models
- ✓ **Commercial usage**

Additional benefits:

- ✓ Custom billing
- ✓ Enterprise features
- ✓ Everything in Professional

https://stability.ai/membership#select_membership

Models



Stability AI Company

https://stability.ai/membership stabilityai Stability-AI

AI & ML interests

Our vibrant communities consist of experts, leaders and partners across the globe. They are developing cutting-edge open AI models for Image, Language, Audio, Video, 3D and Biology.

Team members 76

Collections 6

Video
Stability AI's suite of image-to-video models

- stabilityai/stable-video-diffusion-img2vid
- stabilityai/stable-video-diffusion-img2vid-xt
- stabilityai/stable-video-diffusion-img2vid-xt-1-1
- stabilityai/sv3d

Stable LM
Suite of LLMs trained on English

- stabilityai/stablelm-2-12b-chat
- stabilityai/stablelm-2-12b
- stabilityai/stablelm-2-1_6b-chat
- stabilityai/stablelm-2-1_6b

Expand 6 collections

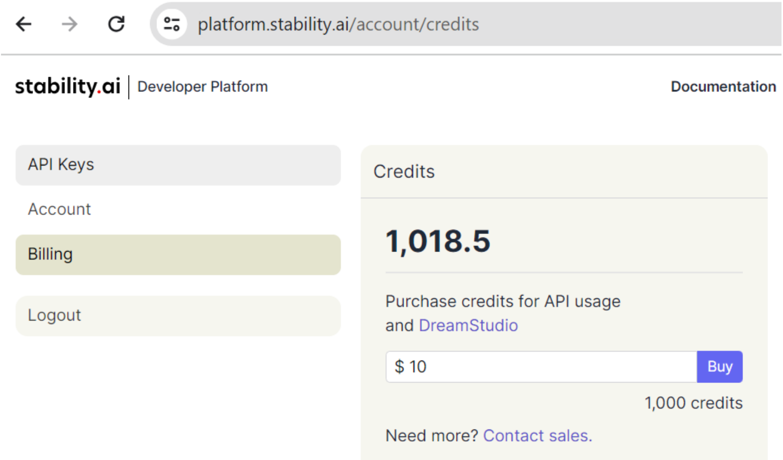
Spaces 6

Sort: Recently updated

- Stable Diffusion 2-1 (pinned, Running on CPU UPGRADE, 10.3k)
- StableLM 2 12B Chat (Running on ZERO, 24)
- Stable Code Instruct 3b (Running, 91)
- TripoSR (Running on ZERO, 608)

https://huggingface.co/stabilityai

API Keys and Credits



platform.stability.ai/account/credits
stability.ai | Developer Platform
Documentation
API Keys
Account
Billing
Logout

Credits

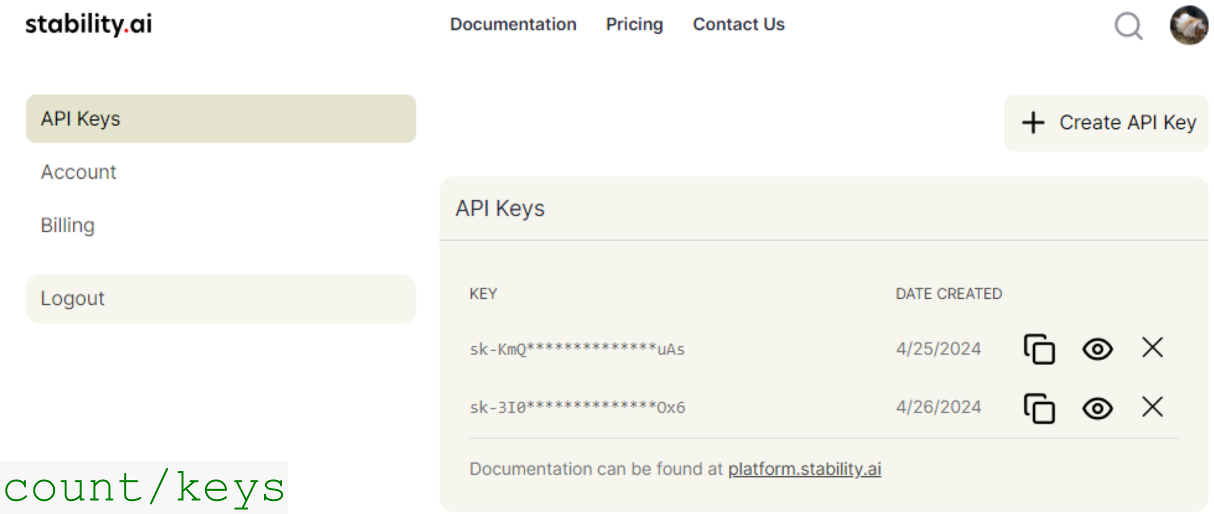
1,018.5

Purchase credits for API usage and DreamStudio

$ 10 Buy

1,000 credits

Need more? Contact sales.



API Keys

KEY	DATE CREATED
sk-KmQ**************uAs	4/25/2024
sk-3I0**************Ox6	4/26/2024

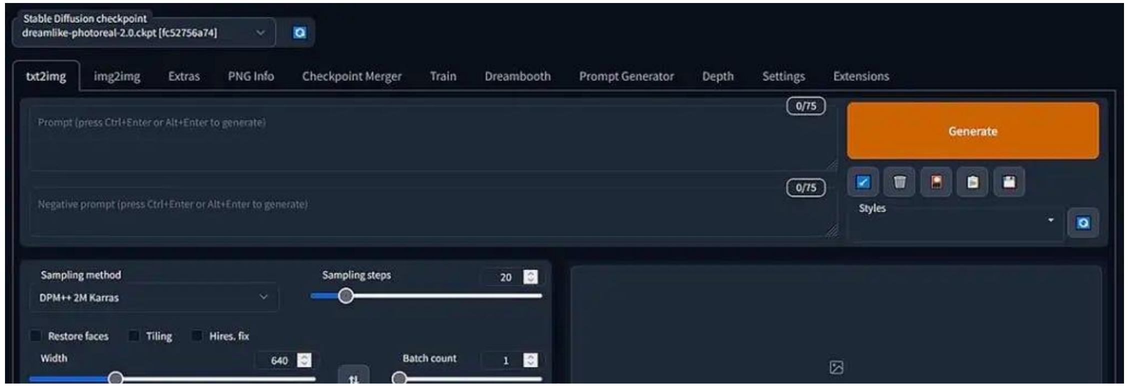
Documentation can be found at platform.stability.ai

https://platform.stability.ai/account/keys

https://platform.stability.ai/account/credits

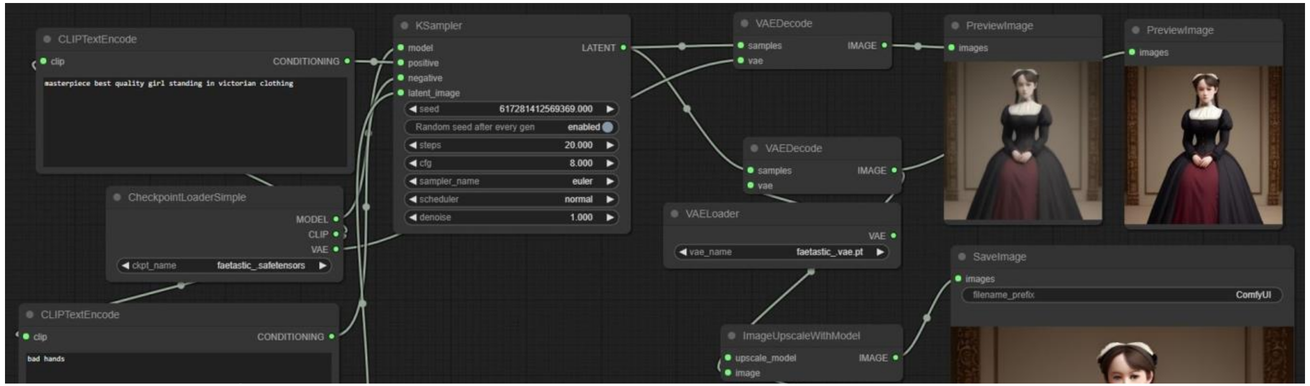
Two Localized Stable Diffusion Apps

Stable Diffusion WebUI



Sampling method: DPM++ 2M Karras
Sampling steps: 20
Width: 640
Batch count: 1

Stable Diffusion ComfyUI



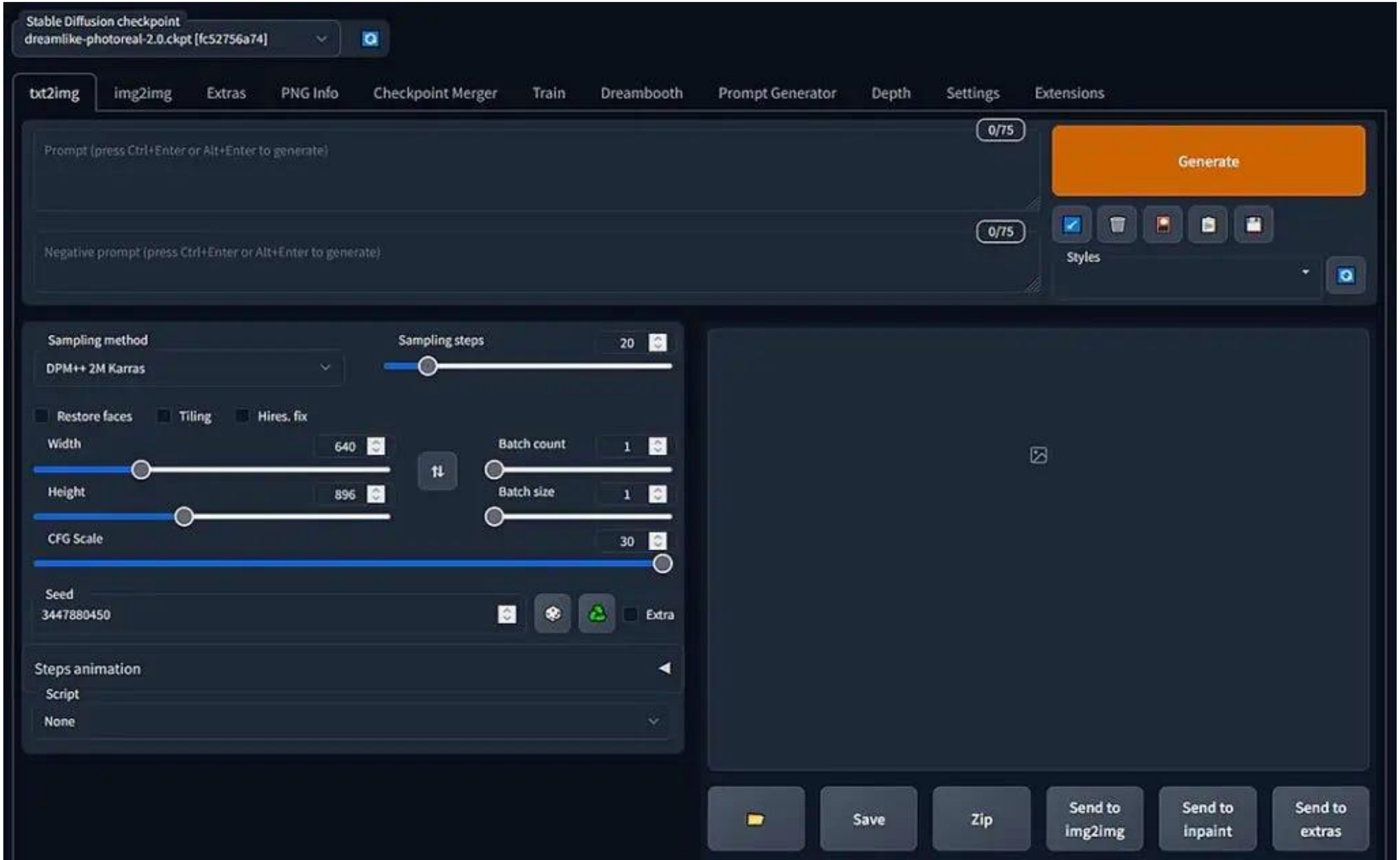
masterpiece best quality girl standing in victorian clothing

bad hands

seed 617281412569369.000
Random seed after every gen enabled
steps 20.000
cfg 8.000
sampler_name euler
scheduler normal
denoise 1.000
ckpt_name faetastic_.safetensors
vae_name faetastic_.vae.pt
filename_prefix ComfyUI

	Stable Diffusion WebUI	Stable Diffusion ComfyUI
Architecture	Gradio	Node-based Layout
UI	Beginner-friendly	More complex
Performance	Relevantly slow	more performant
GPU	>8GB	>4GB require less GPU memory
User control	Less control to the end-user	much greater control to the end-user Easy to repeat 创作者的作品更容易复现
Extensions	More extensions available	Less extensions
Community support	Great support	-
Technical requirement	Entry level	Advanced level
Version	Stable	Can be confused
Recommendation	Normal users	Professional users

Stable Diffusion WebUI



Stable Diffusion checkpoint
dreamlike-photoreal-2.0.ckpt [fc52756a74]

Sampling method
DPM++ 2M Karras

Sampling steps 20

□ Restore faces □ Tiling □ Hires. fix

Width 640

Height 896

Batch count 1

Batch size 1

CFG Scale 30

Seed
3447880450

□ Extra

Steps animation

Script
None

Parameters	Description
Checkpoint	• Stable Diffusion model, a trained model that you need to import to give the model weights. • Two most popular sites for free SD models are huggingface and civitai.
Prompt	• A language representation of what you want the model to generate • **Weight/Attention/Emphasis:** You can adjust the ***weight*** of a word in a prompt using (**word: factor**) where factor is a value greater than zero (<1 means *less important,* while values >1 means *more important)*. For example, we can adjust the weight of the keyword dog in the prompt: "*man and (**dog: 1.8**), playground, rain, trees“*.
Negative Prompt	• Essentially things that you don’t want to appear in your image. • For example: “deformed, blurry image, noise, extra hands”
Sampling Steps	• Parameter to control the number of denoising/diffusion steps. • Usually, higher is better but to a certain degree. The default is 20-30 steps.
Sampling Method	• Algorithm that takes the generated image after each step and compare it to the prompt requested, and then add a few changes to the noise till it gradually reaches an image that matches the prompt description. • Common ones are Euler A, DDIM, and DPM++.
CFG Scale	• Parameter seen as the “Creativity vs Prompt” scale, Classifier-Free Guidance. • Lower number gives the AI more freedom to be creative, while higher number forces it to stick more to the prompt. • The general range of CFG is 5-15. The default CFG 7 gives the best balance between creativity and prompt’s meaning.
Seed	• The randomly generated number which serves as a basis for the image generation process. • Default value -1 means random.
Batch Count **Batch Size**	**Batch Count**: how many batches to generate, one batch will be generated after the other. It doesn’t impact performance. **Batch Size**: how many images to parallelly generate in one batch. ‘1’ is recommended.
Image size	• The image size in pixels. Generating larger images requires more VRAM (GPU memory).
Styles	• Save a prompt as a style for later use.

Tea Break

Hands-on Activity 1

- **Copilot -** https://copilot.microsoft.com/
- **Stable Diffusion WebUI**

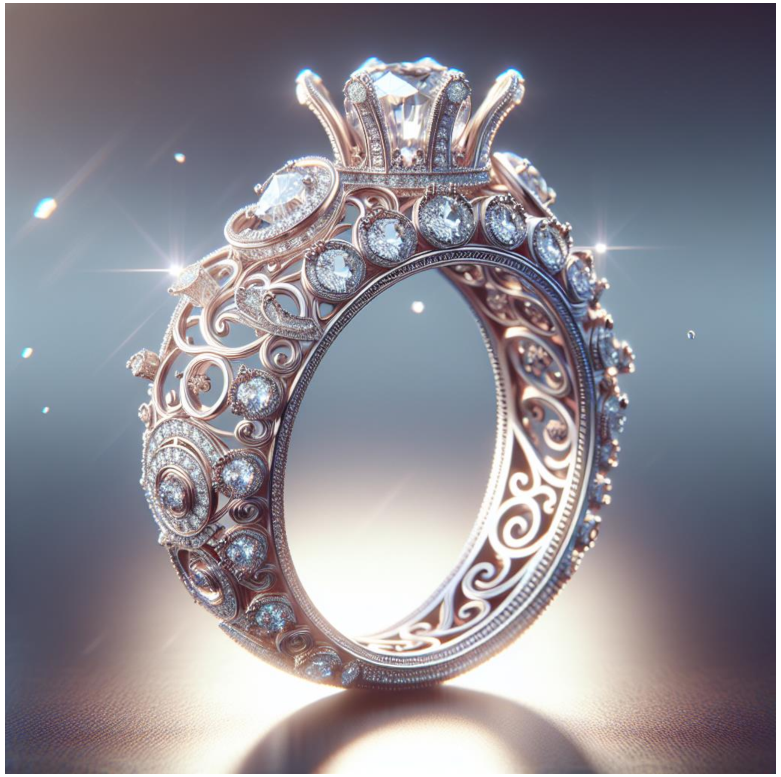
Lunch

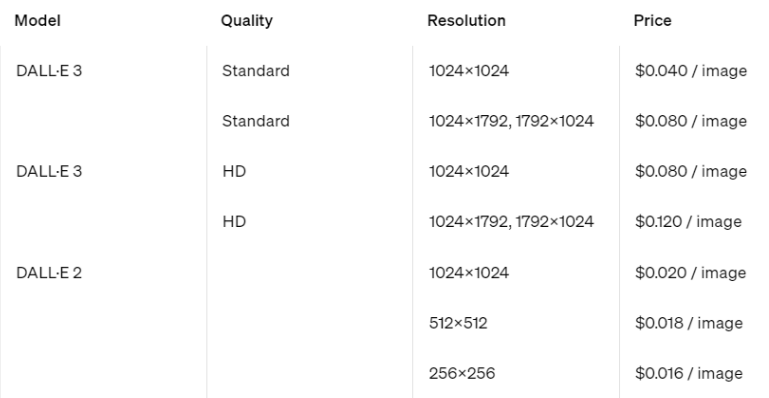
Image Generation using APIs

Image_Generation_Using_API.ipynb

Dall-E through OpenAI API

```
prompt = "a jewelry design, royalty-themed ring, diamonds, 
luxury, closeup, product view, cgsociety, ultra quality, 
digital art, exquisite hyper details,Soft illumination, 
dreamy, fashion, rendering by unreal engine"

# use Dall-E 3 model
response = client.images.generate(
  model="dall-e-3",
  prompt=prompt,
  size="1024x1024",
  quality="standard",
  n=1,
)

print(response.data[0].url)

# To display the image in Google Colab
from IPython.display import Image
import requests

# Fetch the image from the URL
image_data = requests.get(response.data[0].url).content
display(Image(image_data))
```

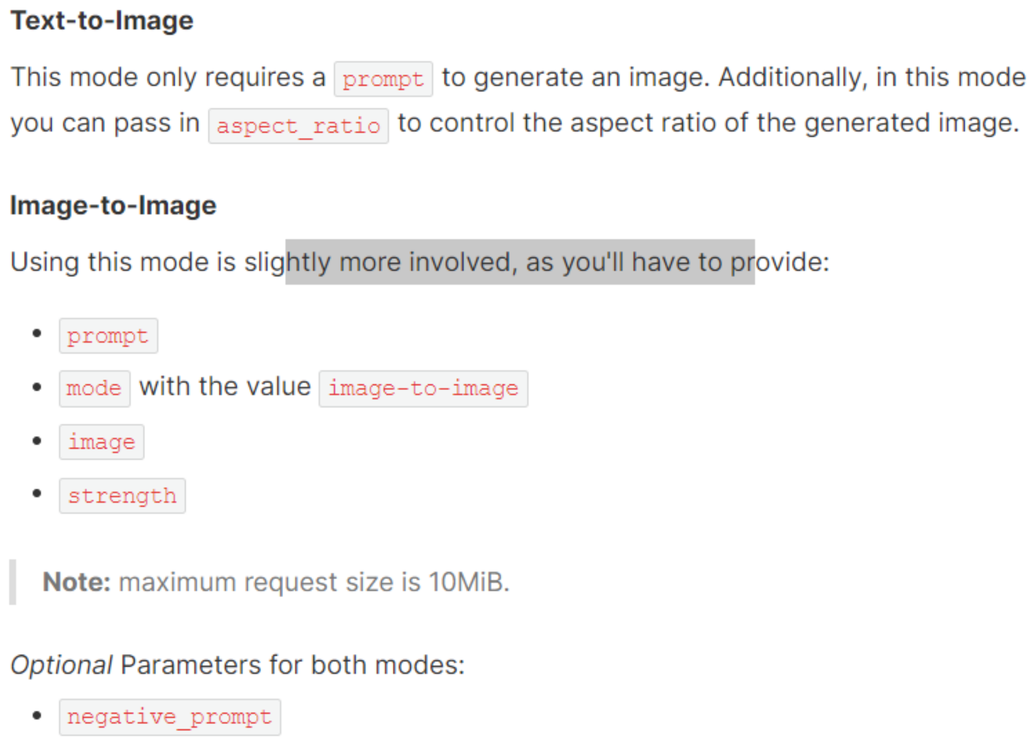




Model	Quality	Resolution	Price
DALL·E 3	Standard	1024×1024	$0.040 / image
	Standard	1024×1792, 1792×1024	$0.080 / image
DALL·E 3	HD	1024×1024	$0.080 / image
	HD	1024×1792, 1792×1024	$0.120 / image
DALL·E 2		1024×1024	$0.020 / image
		512×512	$0.018 / image
		256×256	$0.016 / image

Stable Diffusion API

API Parameters: https://platform.stability.ai/docs/api-reference#tag/Generate/paths/~1v2beta~1stable-image~1generate~1sd3/post



Text-to-Image

This mode only requires a `prompt` to generate an image. Additionally, in this mode you can pass in `aspect_ratio` to control the aspect ratio of the generated image.

Image-to-Image

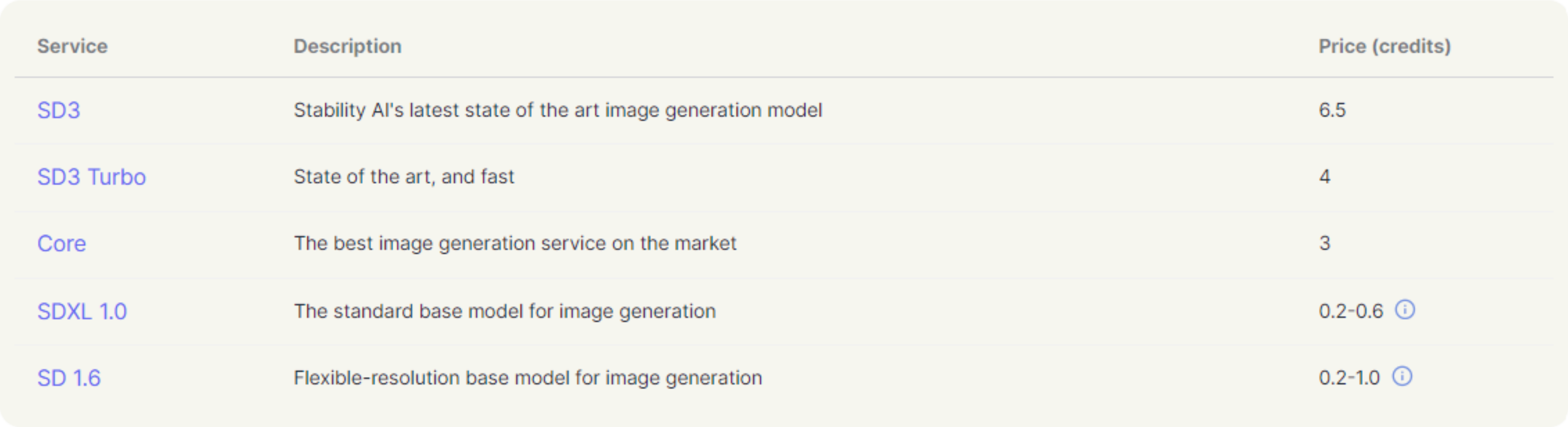
Using this mode is slightly more involved, as you'll have to provide:

- `prompt`
- `mode` with the value `image-to-image`
- `image`
- `strength`

> **Note:** maximum request size is 10MiB.

Optional Parameters for both modes:

- `negative_prompt`



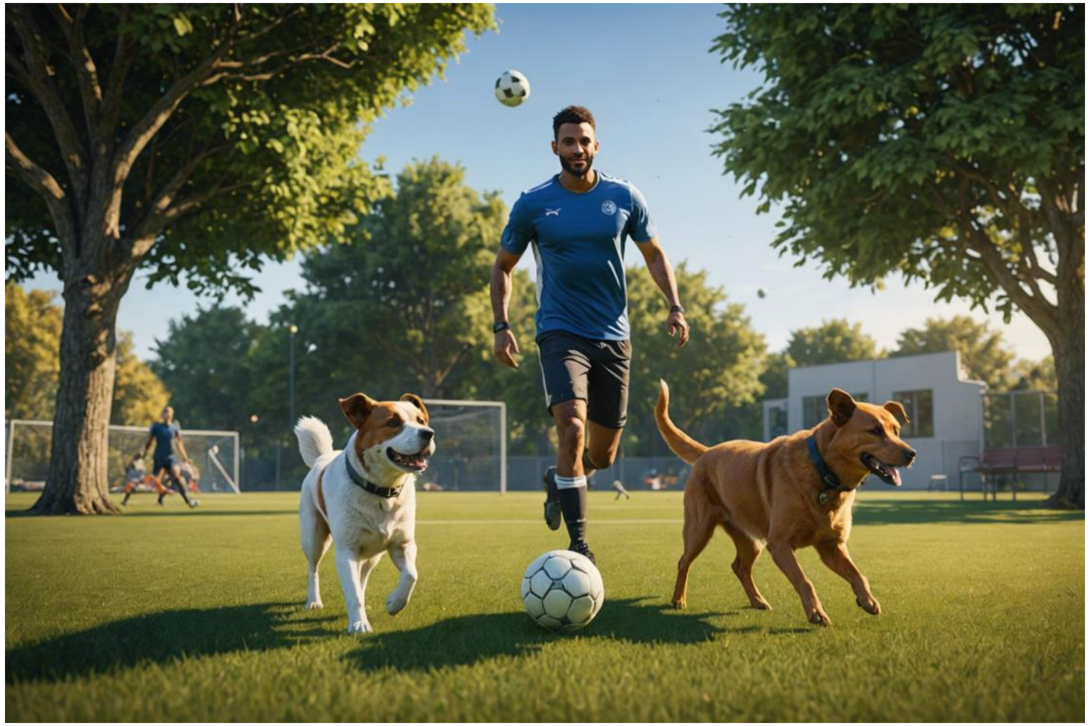
Service	Description	Price (credits)
SD3	Stability AI's latest state of the art image generation model	6.5
SD3 Turbo	State of the art, and fast	4
Core	The best image generation service on the market	3
SDXL 1.0	The standard base model for image generation	0.2-0.6 ⓘ
SD 1.6	Flexible-resolution base model for image generation	0.2-1.0 ⓘ

Stable Diffusion API

- **Generate Image using SD Core model**

```
host = f"https://api.stability.ai/v2beta/stable-
image/generate/core"

params = {
    "prompt" : prompt,
    "negative_prompt" : negative_prompt,
    "aspect_ratio" : aspect_ratio,
    "seed" : seed,
    "output_format": output_format,
    "mode" : "text-to-image"
}

response = send_generation_request(host,params)

# Decode response
output_image = response.content
finish_reason = response.headers.get("finish-reason")
seed = response.headers.get("seed")

# Check for NSFW classification
if finish_reason == 'CONTENT_FILTERED':
    raise Warning("Generation failed NSFW classifier")

# Save and display result
generated = f"generated_{seed}.{output_format}"
with open(generated, "wb") as f:
    f.write(output_image)
print(f"Saved image {generated}")

# Display Image
output.no_vertical_scroll()
print("Result image:")
IPython.display.display(Image.open(generated))
```

```
# Stable Diffusion Parameters
# The default image resolution is 1024x1024.
prompt = "man and (dog:1.8), playing soccer, playground, trees, blue
sky, grass, (sun:0.5), ultra quality, exquisite hyper details,Soft
illumination, rendering by unreal engine"
negative_prompt = "deformed, blurry image, noise, extra hands, extra
feet"
aspect_ratio = "3:2"
seed = 0
output_format = "png"
```



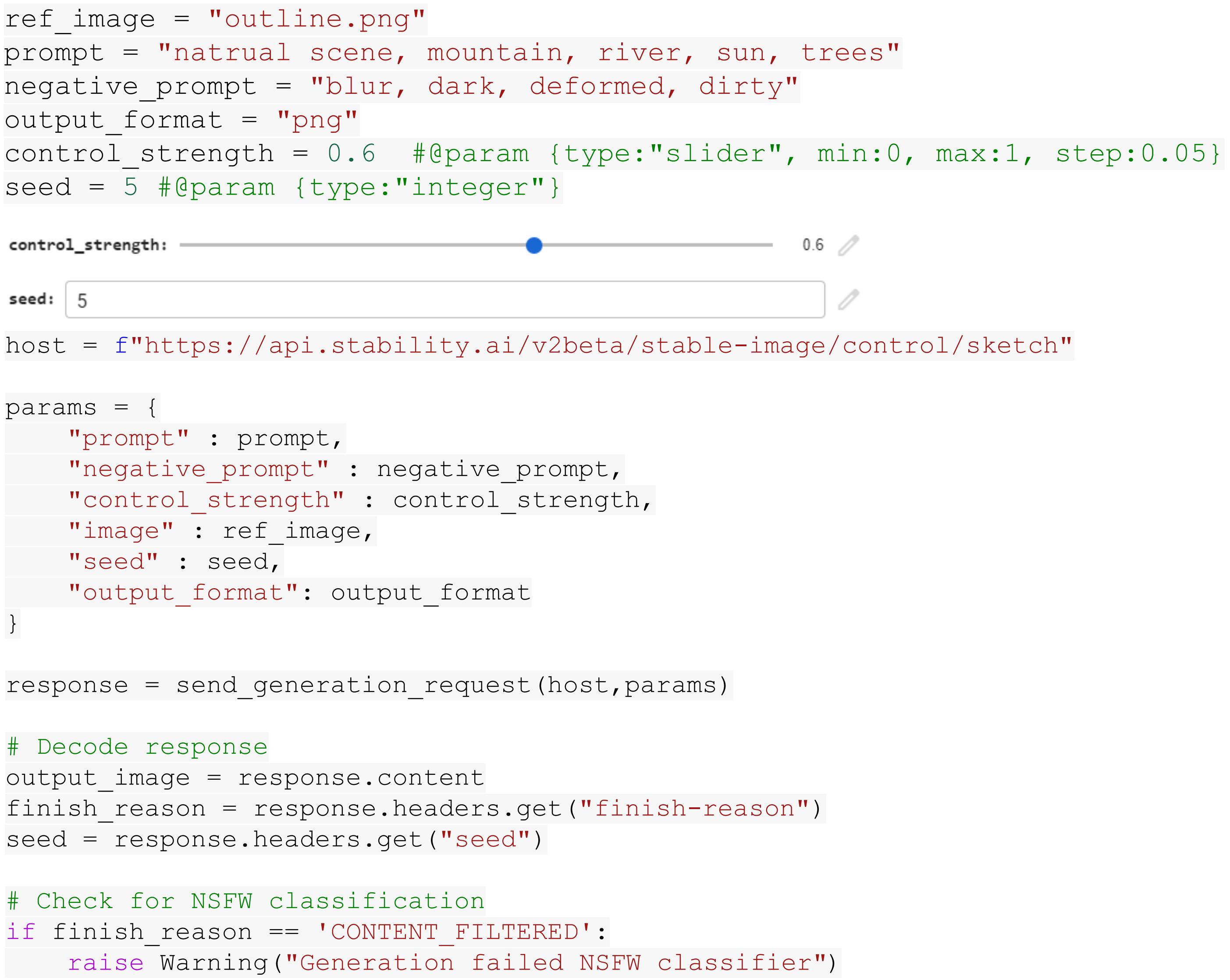
Stable Diffusion API

- **Generate Image using SD3 model**

```
model = "sd3"  #"sd3-turbo"
host = f"https://api.stability.ai/v2beta/stable-image/generate/sd3"

params = {
    "prompt" : prompt,
    "negative_prompt" : negative_prompt if model=="sd3" else "",
    "aspect_ratio" : aspect_ratio,
    "seed" : seed,
    "output_format" : output_format,
    "model" : model,
    "mode" : "text-to-image"
}

response = send_generation_request(host,params)

# Decode response
output_image = response.content
finish_reason = response.headers.get("finish-reason")
seed = response.headers.get("seed")

# Check for NSFW classification
if finish_reason == 'CONTENT_FILTERED':
    raise Warning("Generation failed NSFW classifier")

# Save and display result
generated = f"generated_{seed}.{output_format}"
with open(generated, "wb") as f:
    f.write(output_image)
print(f"Saved image {generated}")

output.no_vertical_scroll()
print("Result image:")
IPython.display.display(Image.open(generated))
```

Stable Diffusion API

- **Generate Image from a Reference Image using SD Sketch model**

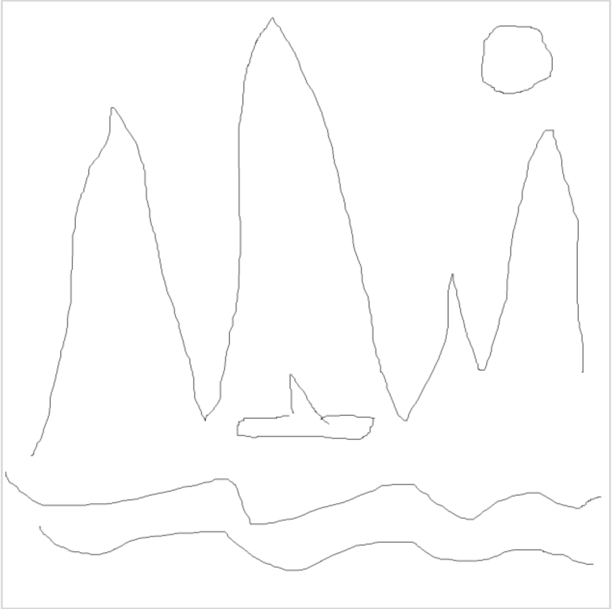


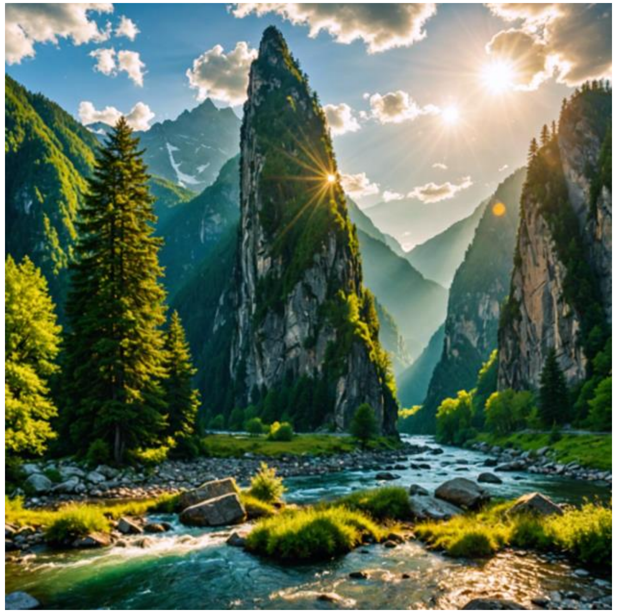
```
ref_image = "outline.png"
prompt = "natrual scene, mountain, river, sun, trees"
negative_prompt = "blur, dark, deformed, dirty"
output_format = "png"
control_strength = 0.6  #@param {type:"slider", min:0, max:1, step:0.05}
seed = 5 #@param {type:"integer"}
```

control_strength: 0.6

seed: 5

```
host = f"https://api.stability.ai/v2beta/stable-image/control/sketch"

params = {
    "prompt" : prompt,
    "negative_prompt" : negative_prompt,
    "control_strength" : control_strength,
    "image" : ref_image,
    "seed" : seed,
    "output_format": output_format
}

response = send_generation_request(host,params)

# Decode response
output_image = response.content
finish_reason = response.headers.get("finish-reason")
seed = response.headers.get("seed")

# Check for NSFW classification
if finish_reason == 'CONTENT_FILTERED':
    raise Warning("Generation failed NSFW classifier")
```





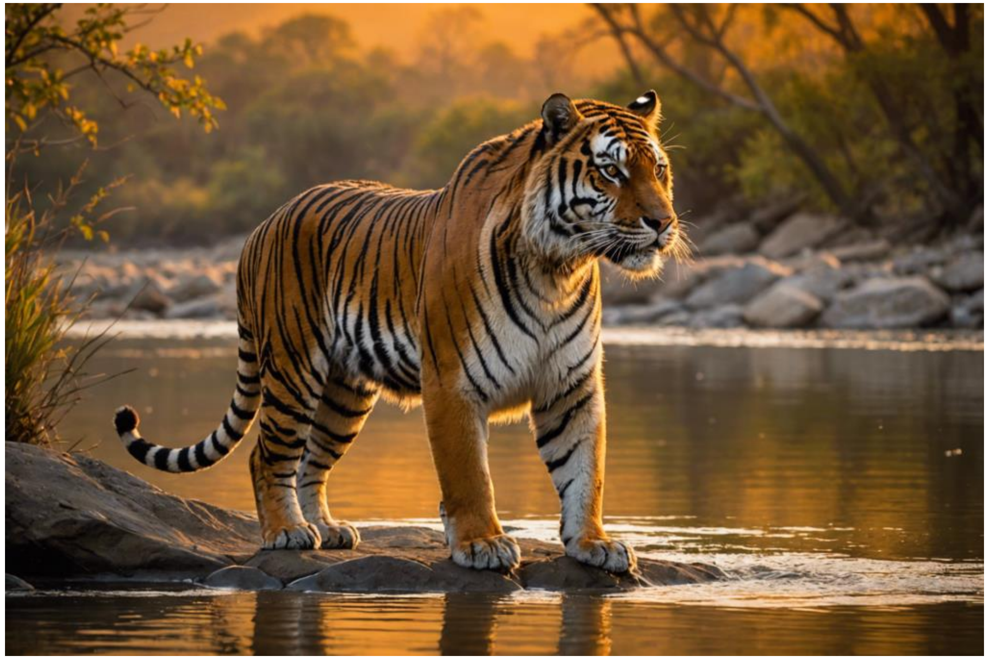
Tea Break

Hands-on Activity 2

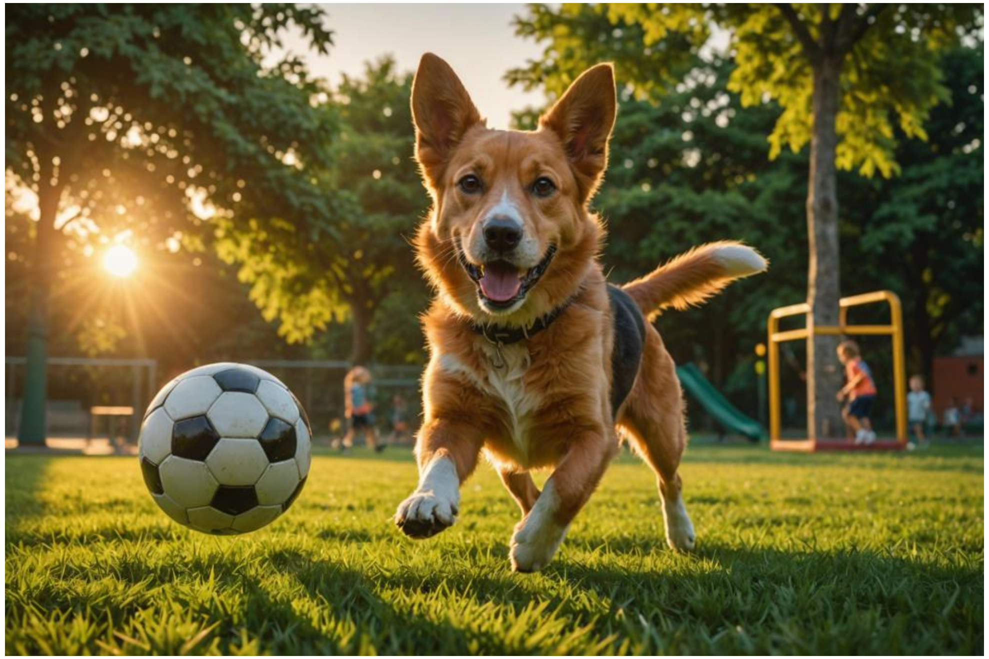
- **Familiar with Image Generation using APIs**
  - Run Image_Generation_Using_API.ipynb with below tasks:
    - Generate **ONE** image using own prompt using Dall-E model
    - Generate **ONE** image using own prompt/negative prompt using SD Core model
    - Generate **ONE** image using own revised outline.png using SD Sketch model
- **Automated and Integrated Prompt Generation (OpenAI API) with Image Generation (Stability.AI API)**
  - Submit notebook Answer

system_message: “You act as an artistic Stable Diffusion prompt assistant. Your task is to generate a detailed, high-quality Stable Diffusion prompt within 100 words. Prompt is used to describe the image, consisting of words separated by commas. The prompt contains the subject of the image, material, additional details, image quality, artistic style, color and lighting. The subject of the image summarizes the main details of the subject (person, thing, scene). For people, you must describe the eyes, nose, and lips, using 'beautiful eyes, lips, extremely detailed face, long eyelashes'. You can also describe the appearance, emotion, clothing, posture, perspective, action, background, etc. Materials used to make artwork using illustration, oil painting, 3D rendering, and photography. Image quality starts with best quality, 4k, ultra-detailed, realistic, photorealistic. Adding artistic styles include: portraits, landscape, anime, photography, concept artists, etc. Adding color tone and lighting effects to control the overall image. Wait for my request to generate prompt.”

user_message:“river, tiger, mountain, trees, sunset”



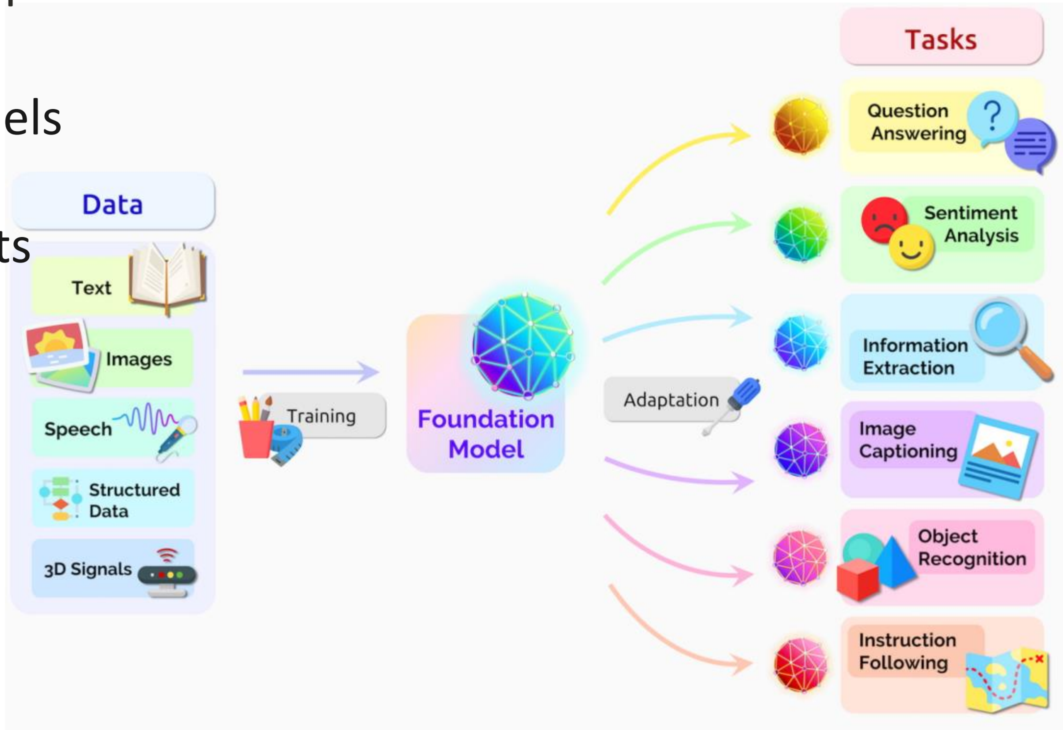
user_message: "playground, dog, soccer, trees, sunset"



Foundation Models

What is Foundation Models?

- Born in Stanford University (Center for Research on Foundation Models (CRFM))
- Pre-trained on Vast Amounts of Data, large AI models
- Self-supervised Learning
- Good generalization allowing zero-shot or few shots learning.
- Fine-tuning and Prompt Engineering (Adaptable)
- Multiple Modalities
- Examples: LLM, BERT, SAM, DINO and etc.



Two models

Segment Anything Model (SAM)

Generative AI - Edit Anything **Model**

Segment Anything Model (SAM)

The **Segment Anything Model (SAM)** by Meta

  - Produces high quality object masks from input prompts such as points or boxes
  - Trained on a dataset of 11 million images and 1.1 billion masks
  - Strong zero-shot performance on a variety of segmentation tasks.

- Application use cases: Healthcare, Autonomous driving,

Ref: GitHub - facebookresearch/segment-anything: The repository provides code fo
inference with the SegmentAnything Model (SAM), links for downloading the traine
checkpoints, and example notebooks that show how to use the model.

Generative AI – Edit Anything

Object Detection Foundation Model

Grounding DINO by IDEA-Research

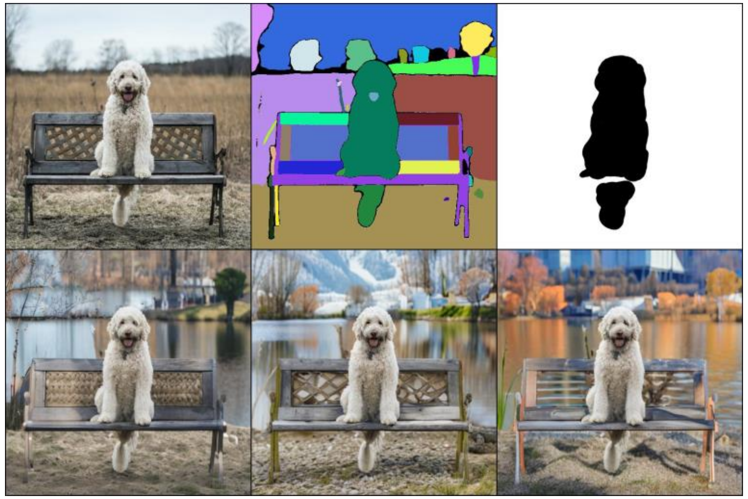
Semantic Segmentation Foundation Model

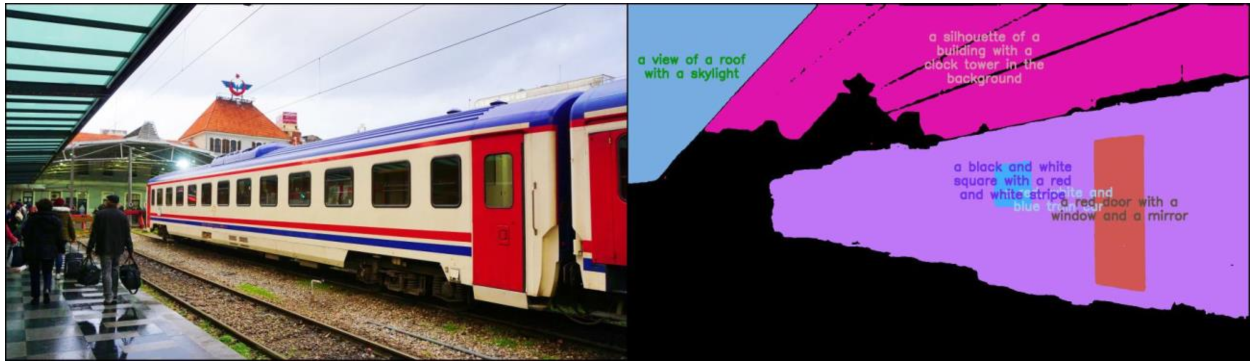
Language-to-Image Generative Model

Image-to-language

BLIP2 by Salesforce

Application use cases: Healthcare, Autonomous driving




a view of a roof with a skylight
a silhouette of a building with a clock tower in the background
a black and white square with a red and white stripe
a red door with a window and a mirror

Ref: GitHub - sail-sg/EditAnything: Edit anything in images powered by segment-anything, ControlNet, StableDiffusion, etc.

Demo

- **Segment Anything Model**
- **Edit Anything Model**

Summary

- Popular Generative AI Platforms for Image Generation
  - Dall-E
  - Midjourney
  - Stable Diffusion
- How Stable Diffusion model works?
- A localized Stable Diffusion WebUI
- Image Generation using APIs
  - Dall-E using OpenAI
  - Stable Diffusion using Stablity.AI
  - Generate Image from prompt
  - Generate Image from sketch
- Integrated Prompt Generation with Image Generation with APIs
  - Prompt Generation with OpenAI + Image Generation with Stablity.AI
- Foundation Models
  - Segment Anything Model
  - Edit Anything Model

References:

- https://en.wikipedia.org/wiki/Stable_Diffusion
- https://gemoo.com/blog/midjourney-vs-stable-diffusion-vs-dalle.htm
- Using Stable Diffusion with webUI in AIME MLC