

# ECE 473: Intro. to Artificial Intelligence

## GenAI Tutorial

Dipesh Tamboli

March 7, 2024

# The Big Picture



- ▶ Overview of GenAI
- ▶ Text to Image generation models
- ▶ Prompt Engineering
- ▶ Finetuning of a Generative Model
- ▶ Generated Image-Prompt alignment score

## Takeaway from the Lecture

1. Should be able to run a text-to-image model
2. Prompt Engineering (designing a better prompt)
3. Score: generated image alignment with the prompt

Source: Image generated locally using the StableDiffusionPipeline from Huggingface.

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ► **Image Generation:**

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation
- ▶ **Text Generation:**

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation
- ▶ **Text Generation:**
  - ▶ Language translation

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation
- ▶ **Text Generation:**
  - ▶ Language translation
  - ▶ Summarization



# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation
- ▶ **Text Generation:**
  - ▶ Language translation
  - ▶ Summarization
- ▶ **Multimodal Generation:**

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation
- ▶ **Text Generation:**
  - ▶ Language translation
  - ▶ Summarization
- ▶ **Multimodal Generation:**
  - ▶ Image captioning

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

- ▶ **Image Generation:**
  - ▶ Realistic image generation
  - ▶ Art generation
- ▶ **Text Generation:**
  - ▶ Language translation
  - ▶ Summarization
- ▶ **Multimodal Generation:**
  - ▶ Image captioning
  - ▶ Visual storytelling

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ▶ **Image Generation:**

- ▶ Realistic image generation
- ▶ Art generation

### ▶ **Text Generation:**

- ▶ Language translation
- ▶ Summarization

### ▶ **Multimodal Generation:**

- ▶ Image captioning
- ▶ Visual storytelling

### ▶ **Video Generation:**

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ▶ **Image Generation:**

- ▶ Realistic image generation
- ▶ Art generation

### ▶ **Text Generation:**

- ▶ Language translation
- ▶ Summarization

### ▶ **Multimodal Generation:**

- ▶ Image captioning
- ▶ Visual storytelling

### ▶ **Video Generation:**

- ▶ Video synthesis

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ▶ **Image Generation:**

- ▶ Realistic image generation
- ▶ Art generation

### ▶ **Text Generation:**

- ▶ Language translation
- ▶ Summarization

### ▶ **Multimodal Generation:**

- ▶ Image captioning
- ▶ Visual storytelling

### ▶ **Video Generation:**

- ▶ Video synthesis

### ▶ **Music Generation:**

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ▶ **Image Generation:**

- ▶ Realistic image generation
- ▶ Art generation

### ▶ **Text Generation:**

- ▶ Language translation
- ▶ Summarization

### ▶ **Multimodal Generation:**

- ▶ Image captioning
- ▶ Visual storytelling

### ▶ **Video Generation:**

- ▶ Video synthesis

### ▶ **Music Generation:**

- ▶ Melody generation

# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ▶ **Image Generation:**

- ▶ Realistic image generation
- ▶ Art generation

### ▶ **Text Generation:**

- ▶ Language translation
- ▶ Summarization

### ▶ **Multimodal Generation:**

- ▶ Image captioning
- ▶ Visual storytelling

### ▶ **Video Generation:**

- ▶ Video synthesis

### ▶ **Music Generation:**

- ▶ Melody generation
- ▶ Music composition



# Introduction to genAI

## What is genAI?

Generative Artificial Intelligence (genAI) refers to a set of AI techniques that focus on generating new content rather than just analyzing or classifying existing data. It encompasses various modalities including:

### ▶ **Image Generation:**

- ▶ Realistic image generation
- ▶ Art generation

### ▶ **Text Generation:**

- ▶ Language translation
- ▶ Summarization

### ▶ **Multimodal Generation:**

- ▶ Image captioning
- ▶ Visual storytelling

### ▶ **Video Generation:**

- ▶ Video synthesis

### ▶ **Music Generation:**

- ▶ Melody generation
- ▶ Music composition
- ▶ Lyrics generation

# Introduction to Text to Image Generation

## What is it?

Text to Image Generation is a type of generative AI technique that aims to generate images based on textual descriptions. This technique involves training deep neural networks to understand the semantics of textual input and generate corresponding images that match the descriptions provided. It has various applications in fields such as art generation, content creation, and visual storytelling.

## Examples

- ▶ **Midjourney**
- ▶ **DALL-E 3**
- ▶ **Stable Diffusion**
- ▶ **Imagen**
- ▶ **Muse**
- ▶ **DreamBooth**

# Stable Diffusion

## What is Stable Diffusion?

Stable Diffusion is a text-to-image generation model based on a type of diffusion model called a latent diffusion model. It is trained to remove noise from images in an iterative process.

### Source

Stable Diffusion: Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. "High-resolution image synthesis with latent diffusion models." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10684-10695. 2022.

DDPM Paper: Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." Advances in neural information processing systems 33 (2020): 6840-6851.


# Text to Image Generation Demo

```
[ ] 1 %pip install -qq -U diffusers datasets transformers accelerate ftfy pyarrow==9.0.0

[ ] 1 import torch
2 from diffusers import StableDiffusionPipeline
3 device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
4 # Check out https://huggingface.co/sd-dreambooth-library for loads of models from the community
5 model_id = "sd-dreambooth-library/mr-potato-head"
6
7 # Load the pipeline
8 pipe = StableDiffusionPipeline.from_pretrained(model_id, torch_dtype=torch.float16).to(
9     device
10 )

▶ 1 prompt = "an abstract oil painting of a green elephant"
2 image = pipe(prompt, num_inference_steps=50, guidance_scale=7.5).images[0]
3 image
```

100% ██████████ 50/50 [00:08<00:00, 5.65it/s]



# Text to Text generation - ChatGPT: Conversational AI Model

## Overview of ChatGPT

Large-scale language model developed by OpenAI for generating human-like responses in conversational settings.

# Text to Text generation - ChatGPT: Conversational AI Model

## Overview of ChatGPT

Large-scale language model developed by OpenAI for generating human-like responses in conversational settings.

## Examples and Use Cases

# Text to Text generation - ChatGPT: Conversational AI Model

## Overview of ChatGPT

Large-scale language model developed by OpenAI for generating human-like responses in conversational settings.

## Examples and Use Cases

- ▶ Well, we all know ;)

# Prompt Engineering: Definition

## Definition

Prompt engineering involves crafting specialized prompts or instructions to guide AI models in generating desired outputs.



# Prompt Engineering: Example

Example: AI Chatbots

# Prompt Engineering: Example

## Example: AI Chatbots

- ▶ User Input: "Where to purchase a shirt."

# Prompt Engineering: Example

## Example: AI Chatbots

- ▶ User Input: "Where to purchase a shirt."
- ▶ Engineered Prompt:  
"You are a sales assistant for a clothing company. A user, based in Alabama, United States, is asking you where to purchase a shirt. Respond with the three nearest store locations that currently stock a shirt."

# Prompt Engineering: Example

## Example: AI Chatbots

- ▶ User Input: "Where to purchase a shirt."
- ▶ Engineered Prompt:  
"You are a sales assistant for a clothing company. A user, based in Alabama, United States, is asking you where to purchase a shirt. Respond with the three nearest store locations that currently stock a shirt."
- ▶ Result: Test it out on a prompt enhanced by ChatGPT and use for Image Generation.

# Prompt Engineering: Example

## Example: AI Chatbots


- ▶ User Input: "Where to purchase a shirt."
- ▶ Engineered Prompt:  
"You are a sales assistant for a clothing company. A user, based in Alabama, United States, is asking you where to purchase a shirt. Respond with the three nearest store locations that currently stock a shirt."
- ▶ Result: Test it out on a prompt enhanced by ChatGPT and use for Image Generation.


Source: AWS. "What is Prompt Engineering?" Retrieved from:  
<https://aws.amazon.com/what-is/prompt-engineering/>

# Prompt: "Elephant"

```
1 prompt = "Elephant"  
2 image = pipe(prompt, num_inference_steps=50, guidance_scale=7.5).images[0]  
3 image
```

[4] ✓ 9.6s

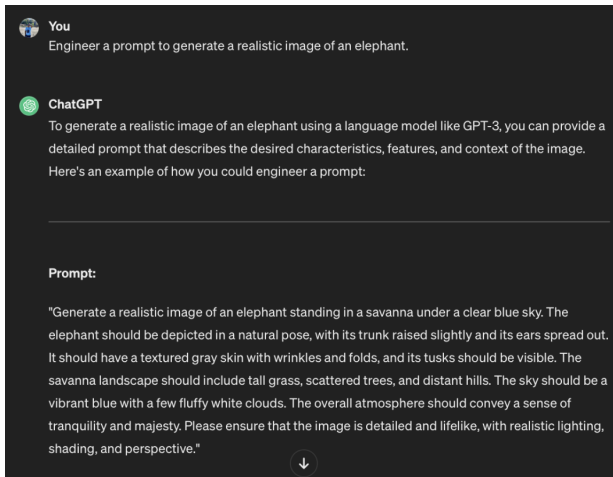
100%  50/50 [00:09<00:00, 5.46it/s]



# Prompt: “realistic image of an elephant”



# ChatGPT engineered prompt



**You**  
Engineer a prompt to generate a realistic image of an elephant.

**ChatGPT**  
To generate a realistic image of an elephant using a language model like GPT-3, you can provide a detailed prompt that describes the desired characteristics, features, and context of the image. Here's an example of how you could engineer a prompt:

---

**Prompt:**

"Generate a realistic image of an elephant standing in a savanna under a clear blue sky. The elephant should be depicted in a natural pose, with its trunk raised slightly and its ears spread out. It should have a textured gray skin with wrinkles and folds, and its tusks should be visible. The savanna landscape should include tall grass, scattered trees, and distant hills. The sky should be a vibrant blue with a few fluffy white clouds. The overall atmosphere should convey a sense of tranquility and majesty. Please ensure that the image is detailed and lifelike, with realistic lighting, shading, and perspective."

↓



```
1 prompt = "Generate a realistic image of an elephant standing in a savanna under a clear blue s
2 image = pipe(prompt, num_inference_steps=50, guidance_scale=7.5).images[0]
3 image
```

[5] ✓ 9.7s

... Token indices sequence length is longer than the specified maximum sequence length for this model (1...  
The following part of your input was truncated because CLIP can only handle sequences up to 77 token

... 100% ██████████ 50/50 [00:09<00:00, 5.41it/s]

...


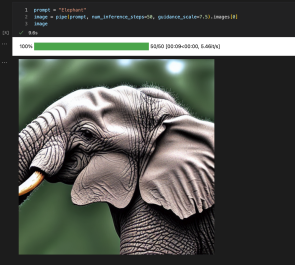


Figure: Using ChatGPT generated prompt

# Comparison

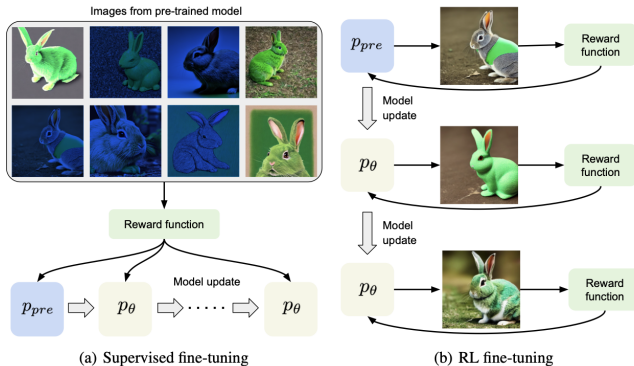


# Finetuning a LLM

## Why?

Fine-tuning involves refining the parameters of a pre-trained large language model to suit a particular task or domain. While pre-trained language models such as GPT boast extensive language understanding, they may lack specialization in specific domains. Fine-tuning addresses this gap by enabling the model to learn from domain-specific data, enhancing its accuracy and effectiveness for targeted applications.

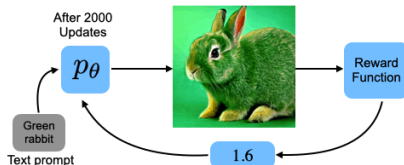
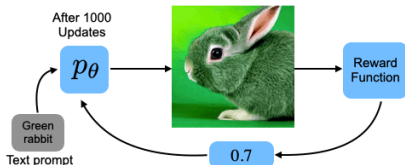
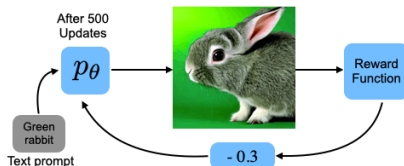
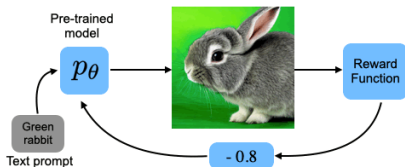
# DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models



## Source

Fan, Ying, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. "Reinforcement learning for fine-tuning text-to-image diffusion models." *Advances in Neural Information Processing Systems* 36 (2024).

# DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models



## Source

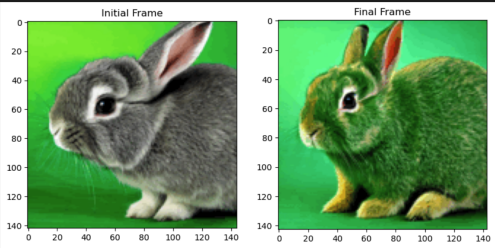
Fan, Ying, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. "Reinforcement learning for fine-tuning text-to-image diffusion models." *Advances in Neural Information Processing Systems* 36 (2024).

# What is ImageReward?

## Demo

```
1 img_init = np.asarray(Image.open("frame_init.png"))
2 img_final = np.asarray(Image.open("frame_final.png"))
3
4 fig, ax = plt.subplots(1, 2, figsize=(10, 5))
5 ax[0].imshow(img_init)
6 ax[0].set_title("Initial Frame")
7 ax[1].imshow(img_final)
8 ax[1].set_title("Final Frame")
9 plt.show()

0.3s
```



```
1 rabbit_prompt = "Green rabbit"
2 reward_init = model.score(rabbit_prompt, "frame_init.png")
3 reward_final = model.score(rabbit_prompt, "frame_final.png")
4
5 print(f"Initial reward: {reward_init}")
6 print(f"Final reward: {reward_final}")

0.1s
```

```
Initial reward: -1.2191205824719238
Final reward: 1.5239769220352173
```

# Resources

- ▶ <https://github.com/Dipeshtamboli/Simple-GenAI-Notebooks>
- ▶ Stable Diffusion: High-resolution image synthesis with latent diffusion models.
- ▶ DDPM Paper: Denoising diffusion probabilistic models.
- ▶ Reinforcement learning for fine-tuning text-to-image diffusion models.
- ▶ AWS. "What is Prompt Engineering?" Retrieved from: <https://aws.amazon.com/what-is/prompt-engineering/>
- ▶ Online Image Text-to-Image <https://deepai.org/machine-learning-model/text2img>
- ▶ <https://generativeai.net>
- ▶ <https://www.techopedia.com/definition/34633/generative-ai>