Building Image Generation Applications

Introduction

In this lesson, we will cover:

- Image generation and why it's useful.
- DALL-E and Midjourney, what they are, and how they work.
- How you would build an image generation app.

Learning Goals

After completing this lesson, you will be able to:

- Build an image generation application.
- Define boundaries for your application with meta prompts.
- Work with DALL-E and Midjourney.

Why build an image generation application?

Image generation applications are a great way to explore the capabilities of Generative AI. They can be used for, for example:

- **Image editing and synthesis**. You can generate images for a variety of use cases, such as image editing and image synthesis.
- **Applied to a variety of industries**. They can also be used to generate images for a variety of industries like Medtech, Tourism, Game development and more.

What is DALL-E and Midjourney?

DALL-E and Midjourney are two of the most popular image generation models, they allow you to use prompts to generate images.

DALL-E

Let's start with DALL-E, which is a Generative AI model that generates images from text descriptions.

> DALL-E is a combination of two models, CLIP and diffused attention.

- **CLIP**, is a model that generates embeddings, which are numerical representations of data, from images and text.
- **Diffused attention**, is a model that generates images from embeddings. DALL-E is trained on a dataset of images and text and can be used to generate images from text descriptions. For example, DALL-E can be used to generate images of a cat in a hat, or a dog with a mohawk.

Midjourney

Midjourney works in a similar way to DALL-E, it generates images from text prompts. Midjourney, can also be used to generate images using prompts like “a cat in a hat”, or a “dog with a mohawk”.

How does DALL-E and Midjourney Work

First, DALL-E. DALL-E is a Generative AI model based on the transformer architecture with an *autoregressive transformer*.

An *autoregressive transformer* defines how a model generates images from text descriptions, it generates one pixel at a time, and then uses the generated pixels to generate the next pixel. Passing through multiple layers in a neural network, until the image is complete.

With this process, DALL-E, controls attributes, objects, characteristics, and more in the image it generates. However, DALL-E 2 and 3 have more control over the generated image.