**Introduction to Generative AI**

**Generative AI and Its Applications**

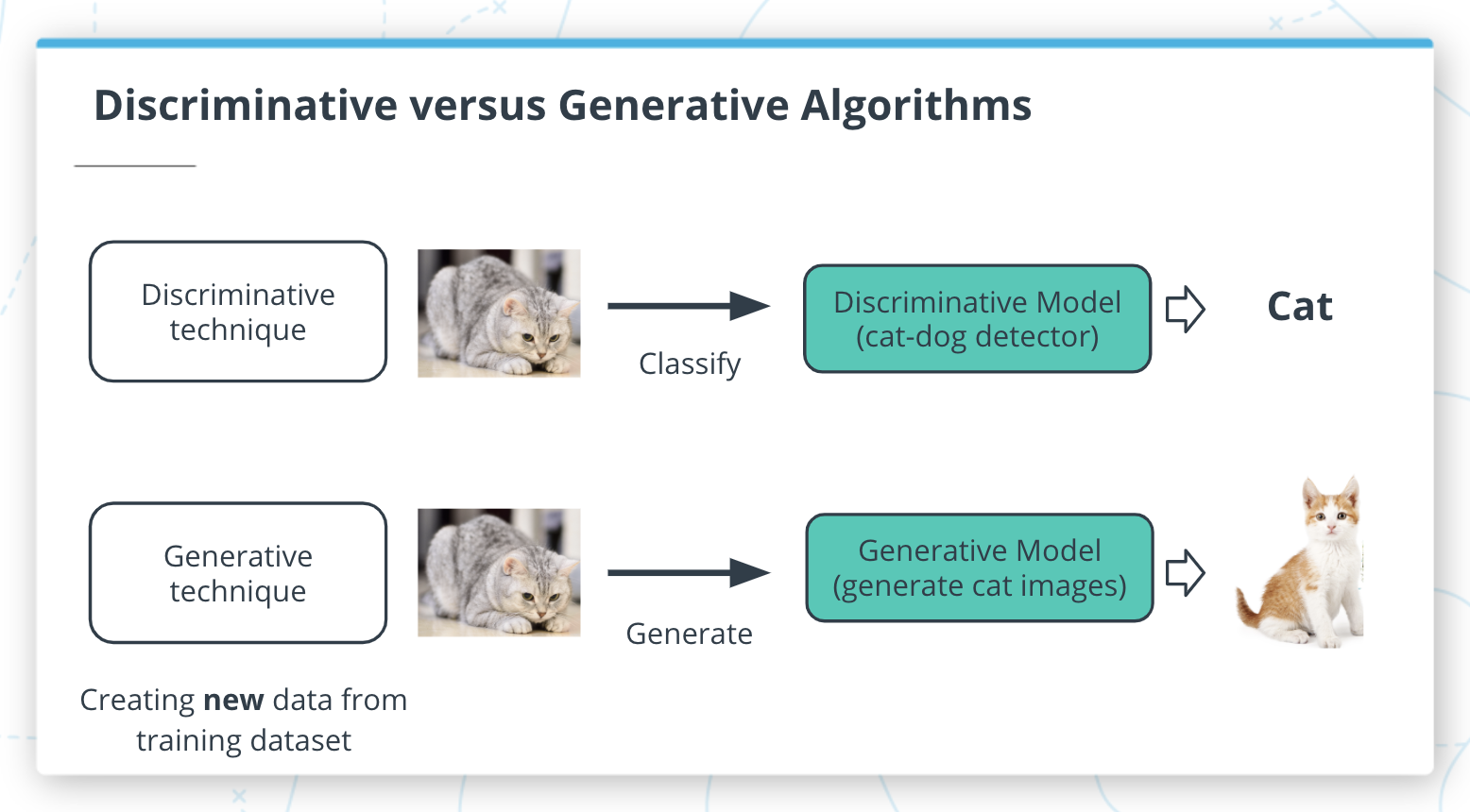
Generative AI is one of the biggest recent advancements in artificial intelligence because of its ability to create new things.

Until recently, the majority of machine learning applications were powered by *discriminative models.* A discriminative model aims to answer the question, "If I'm looking at some data, how can I best classify this data or predict a value?" For example, we could use discriminative models to detect if a camera was pointed at a cat.

As we train this model over a collection of images (some of which contain cats and others which do not), we expect the model to find patterns in images which help make this prediction.

A *generative model* aims to answer the question,"Have I seen data like this before?" In our image classification example, we might still use a generative model by framing the problem in terms of whether an image with the label "cat" is more similar to data you’ve seen before than an image with the label "no cat."

However, generative models can be used to support a second use case. The patterns learned in generative models can be used to create brand new examples of data which look similar to the data it seen before.



Discriminative versus Generative algorithms

**Generative AI Models**

In this lesson, you will learn how to create three popular types of generative models: **generative adversarial networks (GANs), general autoregressive models,** and **transformer-based models**. Each of these is accessible through AWS DeepComposer to give you hands-on experience with using these techniques to generate new examples of music.

**Autoregressive models**

*Autoregressive convolutional neural networks (AR-CNNs)* are used to study systems that evolve over time and assume that the likelihood of some data depends only on what has happened in the past. It’s a useful way of looking at many systems, from weather prediction to stock prediction.

**Generative adversarial networks (GANs)**

*Generative adversarial networks (GANs)*, are a machine learning model format that involves pitting two networks against each other to generate new content. The training algorithm swaps back and forth between training a *generator network* (responsible for producing new data) and a *discriminator network* (responsible for measuring how closely the generator network’s data represents the training dataset).

**Transformer-based models**

*Transformer-based models* are most often used to study data with some sequential structure (such as the sequence of words in a sentence). Transformer-based methods are now a common modern tool for modeling natural language.

We won't cover this approach in this course but you can learn more about transformers and how AWS DeepComposer uses transformers in [AWS DeepComposer learning capsules](https://console.aws.amazon.com/deepcomposer/home?region=us-east-1&utm_source=Udacity&utm_medium=Webpage&utm_campaign=Udacity%20AWS%20ML%20Foundations%20Course#learningCapsules).

;