

**Generative AI with Diffusion Models**

**Course Number:** NVDA-104EC  
**Duration:** 1 day

**Overview**

This NVIDIA Generative AI with Diffusion Models training course teaches attendees how to build a U-Net, a powerful autoencoder designed for images. Participants explore the concept of transposed convolution and discover how it can be used for image manipulation. Students also go beyond the basics and experiment with feeding noise through the U-Net to generate new, unique images.

**Prerequisites**

Attendees must have a good understanding of PyTorch and deep learning.

**Materials**

All attendees receive official courseware from NVIDIA in electronic format.

**Software Needed on Each Student PC**

The class will be conducted in a remote environment that Accelebrate will provide; students will only need a local computer with a web browser and a stable Internet connection. Any recent version of Microsoft Edge, Mozilla Firefox, or Google Chrome will be fine.

**Objectives**

* Build and train a U-Net, a type of autoencoder for images, using transposed convolution to increase the size of an image
* Explore non-sequential neural networks and residual connections
* Experiment with feeding noise through the U-Net to generate new images
* Alter the output of the diffusion process by adding context embeddings and model optimizations
* Understand the CLIP architecture and discover how it associates image embeddings with text embeddings
* Use CLIP to train a text-to-image diffusion model
* Review various state-of-the-art Generative AI models

**Outline**

* Introduction
* From U-Nets to Diffusion
  + Build a U-Net, a type of autoencoder for images
  + Learn about transposed convolution to increase the size of an image
  + Learn about non-sequential neural networks and residual connections
  + Experiment with feeding noise through the U-Net to generate new images
* Control with Context
  + Learn how to alter the output of the diffusion process by adding context embeddings
  + Add additional model optimizations such as
  + Sinusoidal Position Embeddings
  + The GELU activation function
  + Attention
* Text-to-Image with CLIP
  + Walk through the CLIP architecture to learn how it associates image embeddings with text embeddings
  + Use CLIP to train a text-to-image diffusion model
* State-of-the-art Models
  + Review various state-of-the-art Generative AI models and connect them to the concepts learned in class
  + Discuss prompt engineering and how to influence the output of generative AI models better
  + Learn about content authenticity and how to build trustworthy models
* Conclusion