

**Building Conversational AI Applications**

**Course Number:** NVDA-100EC
**Duration:** 1 day

**Overview**

This Building Conversational AI Applications training course teaches attendees how to build AI-powered experiences like virtual assistants, chatbots, and smart speakers. In addition, participants learn how to leverage conversational AI services, including transcription, NLP, and speech, using the NVIDIA® Riva framework.

**Prerequisites**

* Basic Python programming experience
* Fundamental understanding of a deep learning framework, such as TensorFlow, PyTorch, or Keras
* Basic understanding of neural networks

**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**

* Understand automatic speech recognition (ASR) pipelines and critical components
* Craft custom acoustic models, explore decoders, and add punctuation and language identification features
* Uncover the secrets of text-to-speech (TTS) pipelines and customize pronunciations for natural-sounding interactions
* Integrate ASR, NLP (Natural Language Processing), and TTS into a full pipeline and deploy it on the NVIDIA Riva platform
* Optimize performance, tackle scaling issues, and master Kubernetes cluster deployments for real-world applications

**Outline**

* Introduction to Conversational AI
	+ Explore the conversational AI landscape and gain a deeper understanding of the key components of ASR pipelines
	+ Work through an ASR model example from audio to spectrogram to text
	+ Explore decoders, customizations, and additional models, including inverse text normalization (ITN), punctuation and capitalization, and language identification
	+ Deploy Riva ASR
* Customized Conversational AI Pipelines
	+ Explore the key components of the TTS pipeline and full pipeline customizations
	+ Explore the spectrogram generator model and the vocoder model
	+ Work with text normalization and grapheme to phoneme (G2P) conversion to customize pronunciations
	+ Deploy a full ASR-NLP-TTS custom pipeline in Riva
* Inference and Deployment Challenges
	+ Explore challenges related to performance, optimization, and scaling in the production deployment of conversational AI applications
	+ Gain an understanding of the inference deployment process
	+ Analyze non-functional requirements and their implications
	+ Use a Helm chart to deploy a conversational AI application with a Kubernetes cluster
* Conclusion