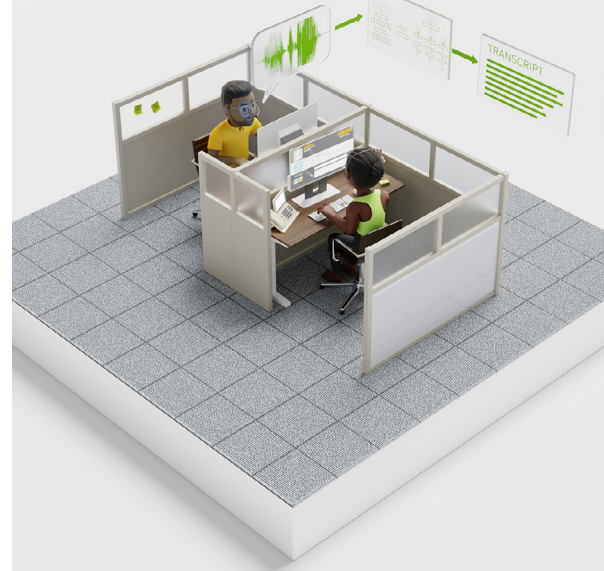




Building Conversational AI Applications

NVIDIA Deep Learning Institute



Workshop Overview

Conversational AI is the technology that powers automated messaging and speech-enabled applications, and its applications are used in diverse industries to improve the overall customer experience and customer service efficiency.

Conversational AI pipelines are complex and expensive to develop from scratch. In this course, you'll learn how to build conversational AI services using the **NVIDIA® Riva framework**. With Riva, developers can create customized language-based

AI services for intelligent virtual assistants, virtual customer service agents, real-time transcription, multi-user diarization, chatbots, and much more.

In this workshop, you'll learn how to quickly build and deploy a conversational AI pipeline including transcription, NLP, and speech. You'll explore automatic speech recognition (ASR) and text-to-speech (TTS) models and their customization in detail with the **NVIDIA NeMo framework** and learn how to deploy the models with Riva. Finally, you'll explore the production-level deployment performance and scaling considerations of Riva services with Helm charts and Kubernetes clusters.

Learning Objectives

In this workshop, you'll learn:

- > How to customize and deploy ASR and TTS models on Riva
- > How to build and deploy an end-to-end conversational AI pipeline, including ASR, NLP, and TTS models, on Riva
- > How to deploy a production-level conversational AI application with a Helm chart for scaling in Kubernetes clusters.

Overview

Duration	8 hours
Price	Contact us for pricing.
Prerequisites	> Basic Python programming experience > Fundamental understanding of a deep learning framework, such as TensorFlow, PyTorch, or Keras > Basic understanding of neural networks
Tools, libraries, and frameworks	NVIDIA Riva, NVIDIA NeMo, Kubernetes
Assessment type	> Skills-based coding assessments evaluate your ability to build a conversational AI application. > Multiple-choice questions evaluate your understanding of the conversational AI concepts presented in the class.
Certificate	Upon successful completion of the assessment, participants will receive an NVIDIA DLI certificate to recognize their subject matter competency and support professional career growth.
Hardware Requirements	Desktop or laptop computer capable of running the latest version of Chrome or Firefox. Each participant will be provided with dedicated access to a fully configured, GPU-accelerated workstation in the cloud.
Language	English

Sample Workshop Outline

Introduction (15 minutes)	Meet the instructor. > Create an account at courses.nvidia.com/join
Introduction to Conversational AI (120 minutes)	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.
Break (60 minutes)	
Customized Conversational AI Pipelines (120 minutes)	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.
Break (15 minutes)	
Inference and Deployment Challenges (120 minutes)	Explore challenges related to performance, optimization, and scaling in 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.
Break (15 minutes)	
Final Review (15 minutes)	> Review key learnings and answer questions. > Complete the assessment and earn a certificate. > Complete the workshop survey. > Learn how to set up your own AI application development environment.
Next Steps	Continue learning with these DLI trainings: > Fundamentals of Deep Learning > Building Transformer-Based Natural Language Processing Applications > Accelerating Data Engineering Pipelines

Why Choose NVIDIA Deep Learning Institute for Hands-On Training?

- > Access workshops from anywhere with just your desktop/laptop and an internet connection. Each participant will have access to a fully configured, GPU-accelerated server in the cloud.
- > Obtain hands-on experience with the most widely used, industry-standard software, tools, and frameworks.
- > Learn to build deep learning and accelerated computing applications for industries, such as healthcare, robotics, manufacturing, accelerated computing, and more.
- > Gain real-world experience through content designed in collaboration with industry leaders, such as the Children's Hospital of Los Angeles, Mayo Clinic, and PwC.
- > Earn an NVIDIA DLI certificate to demonstrate your subject matter competency and support your career growth.

Ready to Get Started?

For the latest DLI workshops and trainings, visit

www.nvidia.com/dli

For questions, contact us at nvdli@nvidia.com