



Rapid Application Development using Large Language Models (LLMs)



Workshop Overview

Recent advancements in both the techniques and accessibility of large language models (LLMs) have opened up unprecedented opportunities for businesses to streamline their operations, decrease expenses, and increase productivity at scale. Enterprises can also use LLM-powered apps to provide innovative and improved services to clients or strengthen customer relationships. For example, enterprises could provide customer support via AI virtual assistants or use sentiment analysis apps to extract valuable customer insights.

In this course, you'll gain a strong understanding and practical knowledge of LLM application development by exploring the open-sourced ecosystem, including pretrained LLMs, that can help you get started quickly developing LLM-based applications.

Learning Objectives

By participating in this workshop, you'll learn how to:

- > Find, pull in, and experiment with the HuggingFace model repository and the associated transformers API
- > Use encoder models for tasks like semantic analysis, embedding, question-answering, and zero-shot classification
- > Use decoder models to generate sequences like code, unbounded answers, and conversations
- > Use state management and composition techniques to guide LLMs for safe, effective, and accurate conversation

Overview

Duration	8 hours
Price	Contact us for pricing.
Prerequisites	Introductory deep learning, with comfort with PyTorch and transfer learning preferred. Content covered by DLI's Getting Started with Deep Learning or Fundamentals of Deep Learning courses or similar experience is sufficient. Intermediate Python experience, including object-oriented programming and libraries. Content covered by Python Tutorial (w3schools.com) or similar experience is sufficient.
Tools, libraries, and frameworks	Python, PyTorch, HuggingFace, transformers, LangChain, LlamaIndex
Assessment type	A programming exercise involving a multi-state dialog manager.
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

Workshop Outline

Introduction

(15 minutes)

Meet the instructor.

- > Create an account at courses.nvidia.com/join

From Deep Learning to Large Language Models

(75 minutes)

Learn how large language models are structured and how to use them:

- > Review deep learning- and class-based reasoning, and see how language modeling falls out of it.
- > Discuss transformer architectures, interfaces, and intuitions, as well as how they scale up and alter to make state-of-the-art LLM solutions.

Break (15 minutes)

Specialized Encoder Models

(45 minutes)

Learn how to look at the different task specifications:

- > Explore cutting-edge HuggingFace encoder models.
- > Use already-tuned models for interesting tasks such as token classification, sequence classification, range prediction, and zero-shot classification.

Break (60 minutes)

Encoder-Decoder Models for Seq2Seq

(75 minutes)

Learn about forecasting LLMs for predicting unbounded sequences:

- > Introduce a decoder component for autoregressive text generation.
- > Discuss cross-attention for sequence-as-context formulations.
- > Discuss general approaches for multi-task, zero-shot reasoning.
- > Introduce multimodal formulation for sequences, and explore some examples.

Decoder Models for Text Generation

(45 minutes)

Learn about decoder-only GPT-style models and how they can be specified and used:

- > Explore when decoder-only is good, and talk about issues with the formation.
- > Discuss model size, special deployment techniques, and considerations.
- > Pull in some large text-generation models, and see how they work.

Break (15 minutes)

Stateful LLMs

(60 minutes)

Learn how to elevate language models above stochastic parrots via context injection:

- > Show off modern LLM composition techniques for history and state management.
- > Discuss retrieval-augmented generation (RAG) for external environment access.

Assessment and Q&A

(60 minutes)

- > Review key learnings.
- > Take a code-based assessment to earn a certificate.

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 nvdl@nvidia.com

© 2023 NVIDIA Corporation. All rights reserved. NVIDIA and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. 2979053. OCT23

