



使用多模态模型构建 AI 智能体

课程概述

就像人类拥有多种感官来感知周围的世界一样，计算机也有各种传感器来帮助感知人类世界。在医疗行业，计算机断层扫描 (CT) 提供了用于检测潜在危险异常的 3D 表示。机器人行业，激光雷达被用来帮助机器人感知深度并在周围复杂的地形中导航。本课程中，学员将探索不同的融合技术，开发能使用多种不同数据类型进行推理的神经网络智体。

学习目标

参加本次培训，您将学到：

- > 不同的数据类型及如何使其适用于神经网络
- > 模型融合，以及早期融合、晚期融合和中期融合的区别
- > 使用 OCR 进行 PDF 提取
- > 模态和智能体编排的区别
- > 使用视频搜索和摘要 (VSS) 定制 NVIDIA AI Blueprints

课程概要

课程时长	8 小时，课后 6 个月内可以继续访问课件和实验 (实验资源用量有限额)
课程模式	讲师授课，每位学员使用完全配置的云端实验环境进行实战练习
预备知识	> 对深度学习概念有基本的理解 > 熟悉至少一个深度学习框架，比如 TensorFlow，PyTorch 或者 Keras，本课程使用的是 PyTorch
使用的工具、库和框架	Python，CLIP
课程评测方式	基于所学实现模型映射
培训证书	成功完成本课程和测试后，学员将获得 NVIDIA DLI 培训证书，证明在相关领域的技能，为职业发展提供证明
课程语言	中文
学习此课程的硬件要求	您需要一台能够上网的笔记本电脑或台式机，且能够运行最新版 Chrome 或 Firefox 浏览器；我们为您提供在云端实验环境的专用访问权限
课程价格	微信添加 DLI 小助手 (微信号 NVIDIALearn)，沟通培训需求

课程大纲

课程介绍

(15 分钟)

- > 讲师介绍
- > 登录 DLI 学习平台

早期和晚期融合

(60分钟)

- > 使用相机和 LiDAR 数据预测物体位置
- > 转换各种数据类型，使其适用于神经网络

中间融合

(60 分钟)

- > 探索有效多模态模型架构背后的理论
- > 训练对比预训练模型
- > 创建向量数据库

休息 (60 分钟)

模型映射 (projection)

(120 分钟)

- > 将 Llama 2 转换为视觉语言模型 (VLM)
- > 使用光学字符识别 (OCR) 工具处理 PDF

休息 (15 分钟)

模型编排

(120 分钟)

- > 使用 Cosmos Nemotron 分析视频
- > 使用 VSS 回答用户关于视频内容的查询
- > 使用 NVIDIA AI Blueprints 进行编排

评估测试和总结

(60 分钟)

- > 回顾所学要点并解答问题
- > 通过映射将预训练模型转换为能够接收不同的数据类型，完成评估并获得证书
- > 填写培训调查表

下一步

学习更多 DLI 相关课程：

- > 构建基于扩散模型的生成式 AI 应用
- > 高效定制大语言模型 (LLM)
- > 构建基于大语言模型 (LLM) 的应用
- > 利用提示工程构建大语言模型 (LLM) 应用
- > 构建大语言模型 RAG 智能体
- > 模型并行 —— 构建和部署大型神经网络

为何选择 NVIDIA 深度学习培训中心 (DLI) 的实战培训？

- > 学习 NVIDIA 与技术专家和行业领导者合作开发的课程，获取全球同步、技术领先和现实可用的专业开发技能和经验。
- > 学习使用行业通用、标准的软件、工具和框架进行端到端的应用开发，能够在广泛的行业中构建基于深度学习、加速计算、图形与仿真和数据科学的应用。
- > 系统化地学习理论，并使用云端完全配置的实验环境同步边练，高效提升实战开发能力。
- > 获得 NVIDIA 全球开发者培训证书，加持专业培训认证，助力职业发展。

准备好开始学习了吗？

查询更多 DLI 课程，请访问 nvidia.cn/training

如有疑问，请通过微信联系 DLI 小助手 (微信号 NVIDIALearn)。

© 2025 NVIDIA Corporation。保留所有权利。NVIDIA、NVIDIA 徽标、Clara、CUDA、DGX、DGX SuperPOD、Index 和 Triton 均为 NVIDIA Corporation 在美国和其他国家 / 地区的商标和 / 或注册商标。其他公司和产品名称可能是其各自关联公司的商标。其他所有商标均为其各自所有者的财产。3734222。2025 年 2 月

