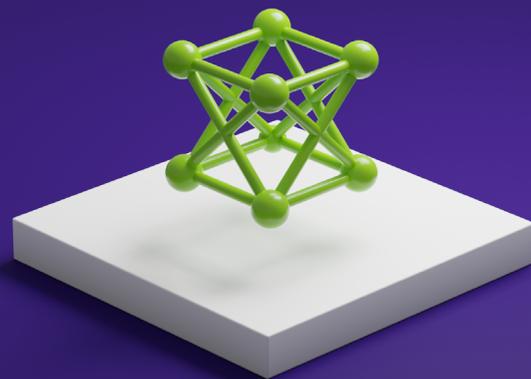




高效定制 大语言模型 (LLM)



课程概述

企业经常需要执行与语言相关的任务，例如文本分类、内容生成、情感分析，或聊天式的客户支持，并寻求以最具性价比的方式来实现这一目标。大语言模型 (LLM) 可以自动执行这些任务，高效的 LLM 定制技术可以增强模型的功能，并缩小企业应用所需模型的规模。

在本课程中，您不仅会学习如何对 LLM 进行提示词工程，还会学到各种技巧，针对您的特定场景高效地定制预训练 LLM 模型，而无需进行计算密集且昂贵的模型预训练或微调模型权重的过程。通过使用 NVIDIA NeMo™ 开源框架，您将学习各种高效参数微调方法，从而定制适合您企业的 LLM。

学习目标

参加本次培训，您将学到：

- > 在有限的的数据上应用各种微调技术，为特定的应用完成模型定制任务
- > 使用 LLM 创建合成数据，用来微调较小的 LLM 以执行所需的任务
- > 利用 NeMo 框架轻松自定义 GPT、LLaMA-2 和 Falcon 等模型

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，其中 GPU 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练习 |
| 预备知识 | <ul style="list-style-type: none"> > 熟悉 Python 编程语言 > 熟悉基本的深度学习技能，如模型架构、训练和推理；推荐学习《深度学习新手入门》或《深度学习基础 —— 理论与实践入门》 > 熟悉基于 Python 的现代深度学习框架，特别是 PyTorch > 熟悉开箱即用的预训练 LLM；推荐学习《使用大语言模型 (LLM) 快速开发应用》 > 熟悉高级提示词工程 |
| 使用的工具、库和框架 | Python、NVIDIA NeMo Service、GPT、LLaMA-2 |
| 学员评测方式 | 根据编码技能进行评估，评价学员高效定制和构建预训练 LLM 模型的能力 |
| 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 |

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。

课程大纲

课程介绍

(15 分钟)

- > 讲师介绍
- > 登录 DLI 学习平台

参数高效微调 (PEFT) 和监督式微调 (SFT)

(115 分钟)

探索语言模型的低秩适配 (LoRA) 和 p-tuning 技巧，以用有限的场数据根据场景定制 LLM:

- > 了解 LoRA、p-tuning 这类 PEFT 技巧在高效微调 LLM 中的作用
- > 学习如何获取和准备用于 PEFT 的数据
- > 对多种 GPT LLM 应用 LoRA 和 p-tuning，并量化的分析微调模型的表现

休息 (60 分钟)

减小模型尺寸的 PEFT

(115 分钟)

学习一组利用更大的微调语言模型生成合成数据的技巧，这些数据可以用来创建能执行相同任务的较小 PEFT 模型:

- > 为常见的合成数据生成任务设计专门的 LLM 功能
- > 使用从较大模型响应中生成的合成数据执行 LoRA，以微调出一个能执行相同任务的较小模型
- > 建立一个良性循环，利用 LLMs 创建合成数据来微调能够创建更多合成数据的较小 LLM

休息 (15 分钟)

使用定制模型

(115 分钟)

练习编写应用代码，将多个微调模型结合在一起:

- > 为情感分析、抽取式问答、文本生成和人格生成，创建多个微调的小型 LLM
- > 学习创建能以您自己的写作风格进行表达的 LLM
- > 将多个小型微调的 LLM 组合成有意义的应用程序代码

评估测试和总结

(45 分钟)

- > 回顾所学要点并解答问题
- > 完成评估并获得证书: 运用课上所学技术，使用多个小型微调 LLM 来实现一个详细分析客户电子邮件并回复的应用
- > 填写培训调查表

下一步

学习更多 DLI 相关课程:

- > **使用大语言模型 (LLM) 快速开发应用**
- > **构建基于扩散模型的生成式 AI 应用**
- > **模型并行 —— 构建和部署大型神经网络**

准备好开始学习了吗?

查询更多 DLI 课程，请访问 nvidia.cn/dli。

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

© 2024 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。3198376。2024 年 3 月

