Fullstack AI Engineer & FDE Training Program

职业转型与能力进阶 · Career Transform & Upgrade

By Parallight Inc, San Mateo, CA.



San Francisco Bay Area

City-level location · Natural Earth

把你的经验，
升级为 AI 时代的职业竞争力。

系统学习、真实项目、持续打磨。
从全栈 AI 系统构建，到客户与业务现场交付。

Build something that really matters.

10周项目集训 | 半年项目陪跑 | 1对1全程项目定制

半年陪跑与项目定制适用于「项目落地与半年陪跑方案」，报名交流时确认

中国与硅谷湾区，正在招聘怎样的 AI 工程师?

从具体招聘样本，看薪资、经验与交付要求。

公开岗位 · 地点	招聘薪资	经验与具体要求
中国 · Agent 全栈工程师 TT 语音 / 趣丸科技 · 广州 [1]	**¥20—40K / 月** 15 薪	3 年以上软件研发；Agent 工具与任务编排、权限、部署与反馈迭代。
中国 · AI FDE 帆软软件 · 杭州 [2]	**¥20—40K / 月** 14 薪	本科，招聘页标注经验不限；深入客户现场，将业务需求转成可落地的 Agent 方案并交付。
湾区 · Fullstack / AI Engineer Poka Labs · 旧金山 [3]	**US$180—240K / 年** 股权另列	岗位标题要求 4+ 年经验；前后端、生产级 Agent、ERP / CRM 集成、客户流程与可靠性。
湾区 · FDE, AI Infra OpenAI · 旧金山 [4]	**US$185—300K / 年** 基本工资；股权另计	5 年以上工程 / 技术交付经验；客户协作、生产部署、可复现评测；可能需最多 50% 出差。

这些岗位重视的，是把 AI 接入实际系统，并对结果持续负责的能力。

来源：[1] 企业招聘 · 牛客 [2] 企业招聘 · BOSS 直聘 [3] Poka Labs · YC [4] OpenAI Careers

核查日期：2026-10-28。K = 千；中国为人民币月薪及所列发薪月数，美国为美元年薪。仅为不同资历的招聘样本，不是行业平均，也不代表培训后薪资或录用承诺；以招聘方最新说明为准。

Fullstack AI Engineer、FDE，与传统 AI 岗位有何不同?

理解能力范围与交付责任，才能判断自己的进阶方向。

比较维度	传统 AI 岗位 以算法 / 建模岗位为例	Fullstack AI Engineer 全栈 AI 工程师	FDE Forward Deployed Engineer · 前线部署工程师
工作重心	数据、算法、训练与模型评测	把模型、Agent、前后端与数据连成完整系统	进入客户现场，解决业务与技术之间的落地问题
主要交付	模型、实验结论、算法组件与指标	可运行、可维护、可持续迭代的 AI 应用	接入现有环境、通过验收并被实际采用的方案
核心能力	统计与建模、数据质量、训练和实验设计	软件架构、Agent 编排、接口、部署、测试与观测	需求澄清、数据权限、系统集成、试点与交付沟通
如何证明价值	实验可复现，模型指标达到目标	功能有效，质量、成本与稳定性可验证	实际任务完成，客户使用，交接与持续运维可执行

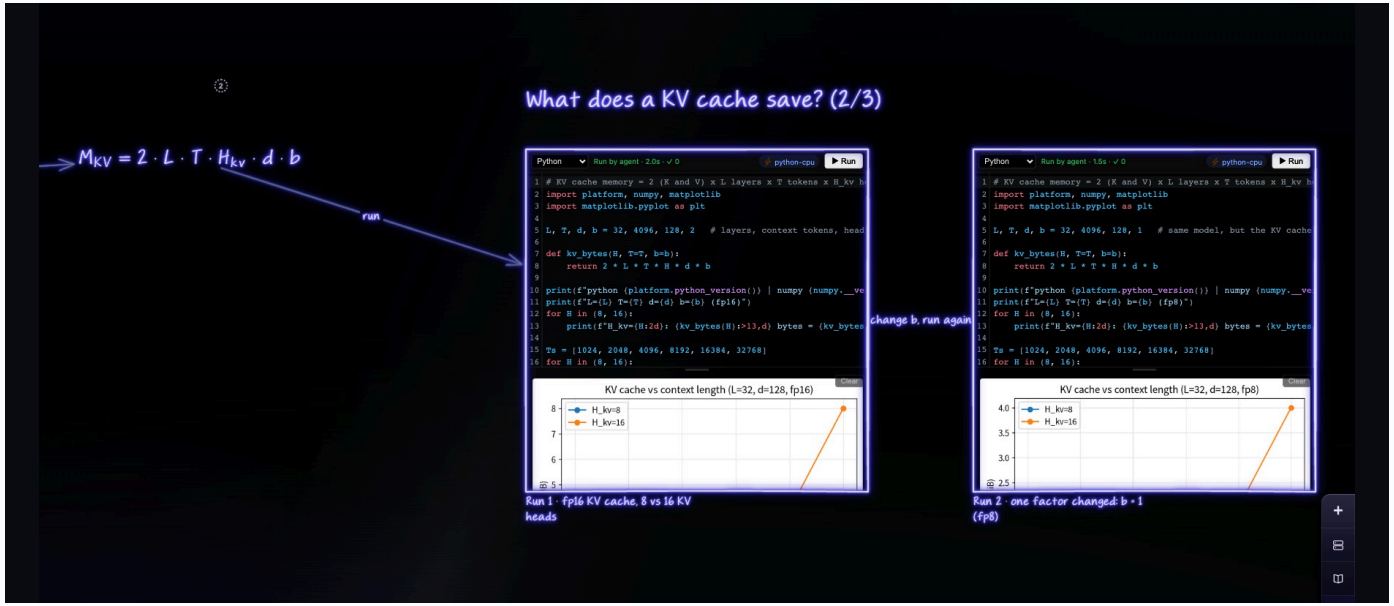
从“我实现了一个模型”，走向“我能把 AI 系统做出来，并让它被用起来”。

上表为依据前页招聘要求整理的能力比较。岗位名称与职责存在交叉；传统 AI 也包含应用、平台和系统岗位，并非仅做算法。两条新路径都需要扎实工程能力，不能只靠工具操作。

我们自己的案例 I：

OpenLabs：全栈 AI 系统实践

OpenLabs：让 AI 讲授、执行实验，并把结果带回学习过程。



现有产品画面 · 概念、代码与运行结果在同一画布上；图中为演示实验。

AI 智能白板

把知识讲解、自动绘图、可运行代码与反馈串起来，让学习者改变条件、重复实验。

自动实验与环境管理

调度 CPU / GPU 与运行环境，执行实验、收集日志和结果，再据此推进下一步。

从项目里带进课堂的方法

DSL、层级结构、纯函数、State Machine：让 Agent 的生成和执行有清晰边界。

上千个

累计管理的运行环境

100+ 门

AI 智能白板讲授的课程

解释 → 执行 → 验证

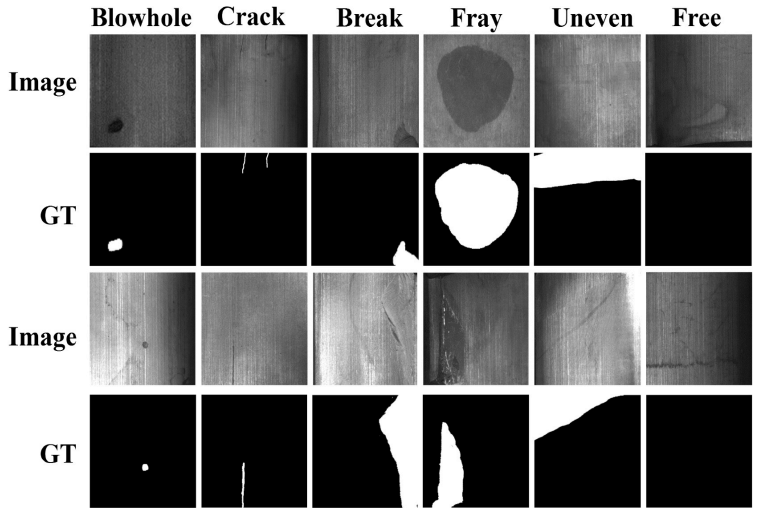
让学习与实验形成闭环

规模信息由项目负责人提供，按累计口径；不表示同时在线规模。产品画面来自 OpenLabs 项目材料。

我们自己的案例 II：

KeploreAI：让工业视觉模型训练自动运行

客户实践 · 帮助某机器视觉设备生产商，构建端到端的全自动模型训练 Agent 系统。



磁性材料 · 磁瓦缺陷检测场景示例

Image 为原图，GT 为人工标注；公开数据集图片，非客户数据或本系统检测结果。

从专业软件操作，到 Agent 驱动的工作流程

自动完成模型训练，释放工程师的工作带宽。

在客户的部分视觉任务中，替代原有 HALCON 工作流程，让工程师减少重复操作，把精力投入问题分析与方案验证。

FDE 的核心实践：把训练方法、执行工具、GPU 环境与设备接口整合成可运行的客户系统。

客户实践与成效由项目负责人提供；HALCON 替代范围限于该客户的部分任务。

系统实现结构

数据与任务 → 自动训练 → 评测反馈 → 模型与设备接口

01 · Skills 构建
沉淀训练方法
任务知识 · 操作规范

→

02 · Tools 优化
封装执行工具
数据处理 · 训练与评测

→

03 · Agent 流程
编排训练任务
步骤 · 状态 · 反馈迭代

→

04 · Runtime 管理
管理运行环境
GPU · 依赖 · 任务生命周期

→

05 · 软硬件整合
连接实际设备
模型接口 · 设备协议

结构图为工程环节示意。配图：Huang、Qiu、Yuan，Magnetic Tile Defect Dataset · “Saliency of magnetic tile surface defects” ↗

导师 | 高老师（Marvin Gao）



高老师

Marvin Gao

Agentist / OpenLabs

教育、学术与 AI 工程

教育经历

毕业于浙江大学和约翰霍普金斯大学计算机科学系。

论文发表

曾在 AAAI、ICRA 等国际会议，以及 TPAMI、JMLR 等学术期刊发表论文。

IBM · AI Scientist

担任项目技术负责人，完成多项企业 B2B AI 项目，包括 PCB 智能化、银行运钞车调度、金融智能客服系统等。

创业与商业化经历

浙江大学期间 · 创业起步

获得浙江大学创业比赛金奖，并获得学校启动资金。

2019 · 企业收购

创立的教育企业被北京一家独角兽企业收购，并成为该企业合伙人。累计培养超过 2,000 名 AI 工程师，分布于全球知名企业。

2025 · KeploreAI 开普勒人工智能

赴美后创办 KeploreAI，获得硅谷投资，投后估值 1,600 万美元。

2026 · Agentist + OpenLabs

在硅谷推进 Agent System 创业，Agentist 与 OpenLabs 累计付费用户超过 500 人。

职业升级的关键：做出经真实场景检验的 AI 系统

从工具使用与局部实现，走向端到端交付：

能构建，能验证，能维护，能持续负责。

Demo

能演示一个想法

真实可用的 AI 系统

稳定运行 · 可维护 · 可扩展

Fullstack AI · FDE

技术系统课 + 产品系统落地课 + 1对1指导



从会做演示，到能交付系统

工程深度 · 真实约束 · 交付证据

课程负责人高老师的独特优势

8年的创业经验 | 10年的 AI工程经验 | 13年的AI科研经验

用持续实践，建立能被看见的职业能力

项目指导方案：第 2 周起，每周至少一次可预约交流

第 2 周 · 定项目

明确问题、使用者与验收目标

项目定义与验收清单

第 8 周 · 首版上线目标

完成核心闭环，供用户试用

部署链接与测试记录

第 10 周 · 验收与迭代

验证可靠性、成本与反馈

验收报告与迭代计划

半年 · 持续使用与迭代

持续验证价值，按目标发展

实验、使用与业务反馈

项目指导适用于 US$1,180 / US$1,880 方案；首版目标一对一约定。半年陪跑仅含于 US$1,880 方案。

把已有经验，转化为 AI 时代的职业竞争力

对希望转入 AI 的专业人士，以及希望扩大影响力的 AI 从业者，这是值得认真投入的职业升级方向。

职业转型 · TRANSFORM

此前从事非 AI 岗位

已有积累：软件开发、数据、产品、业务流程或行业经验，都可以成为理解真实问题的起点。

补齐能力：编程与系统基础、LLM / Agent、工具与数据接入、部署、测试及运行维护。

目标方向：全栈 AI 应用研发，或结合行业经验承担 FDE 场景落地与技术交付。

能力进阶 · UPGRADE

此前已从事 AI 岗位

已有积累：模型、算法、实验与数据经验，为系统的效果判断提供基础。

延伸能力：前后端与服务架构、Agent Runtime、可观测性、成本性能，以及客户系统集成。

目标方向：从局部模块推进到完整系统与场景交付，承担更大范围的技术与业务责任。

共同的进阶产出：一套能接受真实检验的 AI 系统

结合个人背景选题 → 端到端构建 → 测试与部署 → 真实使用反馈 → 持续优化与复盘。

转型路径取决于原有基础、投入与岗位要求。非技术背景需要补足编程和软件工程能力；课程与项目实践不等同于录用、职级或薪资保证。

职业竞争力：用真实系统与个人贡献说话

AI 可以快速生成简历、代码与项目介绍。重复性的项目经历，越来越难体现一个人的能力。

真正有分量的，是你反复打磨、经真实场景检验的 AI 系统。

在真实场景中接受检验

让真实用户完成真实任务。用测试结果、使用反馈与前后对比，说明系统解决了什么问题、带来了什么改变。

在持续迭代中建立深度

把架构、数据、部署与运维做扎实。保留 GitHub 代码、失败记录与优化过程，呈现自己的判断和实际贡献。

形成有辨识度的代表作

结合自己的领域经验，做出独特、可维护的系统。简历与 LinkedIn 指向真实成果，面试中讲清设计取舍与验证过程。

全职集中投入

最快约 2–3 个月

兼顾工作或学业

通常至少半年

越容易“速成”的经历，
越难形成长期竞争力。

这是我们对项目打磨的投入预期，实际进度取决于基础与项目范围。10 周集训建立基础，半年陪跑适用于对应方案；求职结果取决于项目质量与招聘要求。

能力地图：全栈 AI 构建 + FDE 场景交付

10 周技术系统课 + 6 周产品系统落地课，在 10 周集训中组合推进；以真实项目积累工程与交付能力。

技术系统线

AI Agent Native Development

T01—T04 · 多 Agent 协作、MCP、工具与硬件

做出可运行、可部署的原型

Complexity Management

T05—T08 · 状态机、DSL、分层接口与验证

把复杂行为变得可控、可测、可追踪

Performance & Robustness

T09—T10 · 模型服务性能、Token 工厂与运维

让产品稳定运行，并持续优化成本

产品系统落地线

问题识别与系统接入

B01—B02 · 行业判断、真实问题与系统接入

明确接入位置、范围与验收目标

试点验证与使用推广

B03—B04 · 效果与成本、试点与使用推广

用试点验证效果、成本与采用障碍

行业落地与持续交付

B05—B06 · 持续交付、行业问题与应对

形成可维护、可迁移的行业落地方案

3 个系统级项目练习

Enterprise AI Agent / LLM-Server / Embodied AI

提供 GPU、训练环境、数据集、模型、自动测评系统及项目指导。

持续 6 个月的 1 对 1 定制项目陪跑

围绕你希望构建的项目，推进设计、开发、上线与迭代。

适用于项目落地与半年陪跑方案。

课程结构：技术系统课 + 产品系统落地课

T 系列建立全栈 AI 工程能力；B 系列训练 FDE 需要的问题识别、系统接入、试点验证与持续交付。两条线分开介绍。

第一部分

T 系列
技术系统课

T01—T04：Agent 开发、工具协作与硬件

T05—T08：状态机、DSL、系统设计与验证

T09—T10：模型服务、Token 工厂与运维

构建能运行、可验证、可维护的 AI 系统

第二部分

B 系列
产品系统落地课

B01—B02：行业机会、真实问题与已有系统接入

B03—B04：效果成本、试点反馈与使用推广

B05—B06：持续交付、行业典型问题与应对方案

把系统接入真实场景，验证效果并持续改进

完整课纲与学习资源 · 9 次项目构建练习 · 20 门选修课程精选

AI Agent Native Development

技术系统课 · T01—T04

阶段产出

能够组织多 Agent 协作，接入模型、工具与 MCP，将一个想法或开源项目改造成可运行、可部署的 AI 应用，并将 AI 控制延伸到硬件或仿真环境。

- 修改并上线带自定义功能的 VS Code 在线版本，通过 MCP 调用部署工具。
- 构建本地 Agent + MCP Server 的视频合成、剪辑与发布流程。
- 完成可自主学习的智能小车实践，采用实体或仿真方式。

T01 · 研发范式的改变：用 AI 构建与运营复杂产品

本节内容

- 从使用 AI 编写代码，到组织 Agent 完成产品研发、部署、运维、市场调研与用户分析：以 OpenLabs 为案例，展示一个复杂系统如何被拆解、推进并持续迭代。
- 示例：用 AI 动态管理服务器、数据库与 GPU 的配置、部署和运行；从任务需求出发调动资源，读取日志、处理故障并验证结果，展示从“回答问题”到实际执行的完整闭环。
- 从第一性原理理解 Agent 系统：当前 AI 能完成什么、不能完成什么，哪些任务容易、哪些任务难，能力边界与难度来自哪里；为什么演示成功不等于稳定交付。
- Jev 与 World Model 的实现原理与应用分析：理解模型结构、训练目标与推理机制，结合 NanoJev 分析快速决策，讨论世界模型如何用于状态预测与规划，以及与 Agent 的结合方式。
- 这种研发范式如何影响工作方式、岗位需求与产品研发？讨论从单项职责到完整交付的角色变化，以及人与 Agent 的分工和协作。
- 如何在这次技术变革中建立技术优势？结合复杂产品的开发过程，识别仍需积累的工程判断、任务组织与验证能力，持续提升自身价值。

课后实践案例

基于开源 VS Code 的在线版本，用 AI 理解现有代码、运行开发环境，并增加至少一个自定义功能；记录任务拆解、修改过程与功能验证结果，为下一节组织多 Agent 协作与接入自动化部署做准备。

T02 · 面向 Agent 的开发：任务组织、MCP 与多 Agent 协作

本节内容

- 如何选择开发起点？判断哪些项目适合二次开发、哪些需要从头开始；根据改造范围和维护成本确定任务边界。
- 什么是面向 Agent 的开发？将软件能力设计成 AI 可以发现、调用和检查的接口，让 Agent 成为产品的实际使用者。
- MCP、Skills 与 Agent-to-Agent 协议如何分工？分别理解工具接入、可复用操作指引与 Agent 间协作，避免混用概念。
- 如何用 Specification 组织开发？表达需求、拆分任务、约定接口与验收标准，让 Agent 知道做什么、如何判断完成。
- 如何高效实现自己的 MCP Server 和 Tool？设计工具职责、输入输出、调用反馈与错误处理，将已有程序能力开放给 Agent。
- Routine、Hooks 与云端 Project 如何管理持续任务？组织项目上下文、事件触发和可重复执行的流程。
- 多 Agent 如何实际协作？组织 Codex、Claude 与多个 session 并行开发，交接上下文、协调任务依赖、检查并集成成果；练习权限边界、冲突处理与失败恢复。

课后实践案例

继续完善 T01 的在线 VS Code 项目：编写功能与验收 Specification，实现 MCP Server / Tool，组织多个 Agent 开发、检查与集成；配置一个 Hook 或 Routine 自动触发检查，通过 MCP 接入部署工具并上线。提交部署链接和一次完整的执行记录。

T03 · 现代 Agent、AI 工具与能力全景概览

本节内容

- 主流模型与服务如何选型？以 Claude Fable、Claude Opus、GPT Extra、GPT 6.1 Soul、DeepSeek、Gemini、Kimi、GLM 为例，比较适用任务、能力边界、成本与接入方式。
- 什么是决策模型？以 Jev 为例，理解如何使用决策模型，并把它接入产品的决策流程。
- 多模态能力如何组合？梳理图像生成、视频生成与合成、语音合成、自动作曲和剪辑工具，以社交媒体视频制作为例串联完整流程。
- 如何让 AI 获得并操作外部信息？结合浏览器自动操作、API、订阅与定时采集，持续获取数据并保持信息更新。
- World Model、Physical AI 在解决什么问题？理解其核心用途、应用边界，以及与生成内容、执行工具和控制物理设备的区别。
- 如何把能力选型变成产品方案？分析现代 AI 工具带来的技术产品变化，按任务、效果、成本和接入条件组合模型与服务，明确哪些能力直接接入、哪些部分需要自己实现。

课后实践案例

沿用 T02 的 MCP / Tool 设计方法，自建媒体 MCP Server，接入素材或外部信息获取、视频合成、剪辑和发布能力；由本地 Agent 调用服务端完成社交媒体视频工作流。提交可运行演示、工具与模型选型说明，并明确本地端与服务端分工。

T04 · AI 控制硬件

本节内容

- 嵌入式与板卡开发的基本原理：以 Arduino、树莓派为例，理解板卡如何运行程序、连接外设并与外部系统通信。
- 传感器与执行器的工作原理：如何采集环境信息，以及如何驱动电机等设备执行动作。
- AI 控制硬件的两种路径：由 AI 编写控制代码并运行；或让 AI 模型根据观测进行决策，再由控制程序执行动作。
- 复杂场景如何仿真：介绍 ROS 与 Gazebo，了解机器人系统的组织方式，以及仿真环境中的传感器、执行器与任务。
- 如何让 AI 配置、启动、控制和管理 ROS 与 Gazebo，运行实验、读取反馈并调整方案。

课后实践案例

让 AI 完成一辆可以自主学习的智能小车。具体学习任务、学习方法，以及采用实体或仿真的方式，后续共同确定。

Complexity Management

技术系统课 · T05—T08

阶段产出

能够将复杂、难以控制的 AI 程序重构为有状态约束、清晰接口和验证机制的系统，让人能够监督 Agent、定位问题，并用测试证据判断输出是否可信。

- 用状态机优化 AI 销售流程，用自定义 DSL 控制视频、3D 或游戏生成。
- 以 OpenLabs 为例完成分层重构，加入权限、可观察反馈与故障隔离。
- 提交不变量、自动化测试与模拟环境验证结果，说明通过条件与尚未覆盖的边界。

T05 · 数据结构与操作系统思想：优化 AI Native 开发架构

本节内容

- State Machine 状态机如何管理复杂逻辑与复杂 AI 逻辑？以 AI 销售、AI 助理为例，组织状态、事件与任务流程。
- 如何通过不变量（Invariant）管理复杂程序？什么是纯函数、什么是副作用？以 AI 绘图程序为例，让 AI 重构代码，消除不必要的副作用，将必要的输入输出与核心逻辑分离。
- 前缀、共享结构、递归与分形分别起什么作用？讨论它们如何帮助组织序列任务与流式任务，以及适用边界。
- 如何选择搜索方案？比较 RAG、动态 Memory 管理与动态 Context 管理的区别、配合方式及适用场景。

课后实践案例

优化一个 AI 销售的源代码：通过状态机约束其行为，使销售流程更加可控，并支持人类监督、纠正与指导学习。

T06 · 如何通过自建程序设计语言管理复杂逻辑与复杂任务

本节内容

- 什么是领域专用语言（Domain-Specific Language，DSL）？如何用面向特定领域的表达方式描述任务与逻辑。
- 自定义程序设计语言在控制复杂任务中的优势和意义：明确表达规则、约束执行方式，并连接任务描述与实际执行。
- 比较现有的各种 DSL：它们各自的优点、特点、不足与适用范围。
- 如何借助 AI 快速、高效地设计并实现自己的程序设计语言，包括语法、语义、解析、校验与执行。
- 以 AI 自动 PPT 创作、AI 自动绘图和自动游戏构建程序为例，分析 DSL 在其中的作用与价值。

课后实践案例

构建并优化一个 AI 视频、3D 场景或网页游戏生成程序（任选一个方向）：自定义领域专用语言（Domain-Specific Language，DSL），让 AI 先生成描述场景、对象、动画与交互规则的中间表达，再由程序解析、校验并交给渲染引擎执行。通过修改中间表达实现局部调整、重复渲染与行为检查，简化生成过程，提升对生成结果与执行过程的控制力。

T07 · Agent 驱动的复杂 AI 系统的 System Design

本节内容

- 基于实际案例，分析软件架构设计在 Agent 开发中的作用与意义：如何通过模块化、依赖关系设计与系统质量保证，避免局部功能不断堆叠、整体系统逐渐失控。
- 以自动绘图、文档书写与解析程序为例，如何通过重构为多层结构优化程序性能，管理系统复杂性。
- 类似 TCP/IP 的分层协议，对开发复杂系统、管理复杂任务有什么意义和借鉴价值？如何划分层次、定义接口并约束跨层调用。
- 什么时候该重构，如何高效重构？讨论 Agent 驱动的开发为什么更需要持续重构，以及重构范围、渐进迁移、行为保持、回归检查和回滚等注意事项。
- 如何给 AI 增加可观察状态、可反馈信息与可复现条件？结合 Langfuse、自建 AI 调用看板和 AI 状态机可视化三个示例。
- 如何管理 AI 可执行操作的权限？明确工具、数据和环境的访问边界，为关键操作设置必要的授权与审计。
- 如何管理多 Agent 的上下文与任务依赖？协调并行任务、解决修改冲突、集成成果，避免各自完成的局部实现破坏整体架构。

背景引言

代码产生更容易了，软件研发也更容易失控了。如何避免研发烂尾、复杂度无法管理？本节从 System Design 出发，把架构、重构、反馈与协作组织成可持续推进的研发体系。

课后实践案例

以 OpenLabs 自动实验学习环境为例，分析高耦合代码的模块与依赖，制定并实施一次分层重构。明确接口、权限和故障隔离边界，补充可观察状态与回归检查；提交重构前后架构图、依赖关系、可运行代码和验证记录，说明如何定位并阻断异常，以及如何保持原有功能。

T08 · Agent 系统的验证：如何解决 Agent 的幻觉问题，如何相信自己 Agent 产出的东西

本节内容

- 什么是不变量（Invariant）？如何引入并检查不变量，约束程序执行的正确性；区分事实幻觉、逻辑错误、工具失败与业务违例，用来源、测试和实际状态核验。
- 黑盒与白盒测试：比较外部行为验证与内部实现检查，理解各自的作用、覆盖范围与边界。
- 端到端自动化与模拟用户测试：通过实际界面执行完整流程，覆盖异常路径，检查可观察结果并记录可复现的失败轨迹。
- 形式化验证：用 Lean 证明小型程序性质与不变量，检查规格和假设；局部证明不代表整个产品正确。
- 现实验证：用真实环境、传感器或独立测量检查执行结果，对比仿真与实测，明确误差和适用边界。
- 模拟环境：如何定义任务、状态、动作与反馈，建立有效的模拟环境；通过真实数据校准、扰动与边界测试减少 Sim-to-Real Gap。

课后实践案例

设计独立验收：完成一个 Lean 小型证明，构建模拟用户操作的端到端黑盒测试，并对照程序声明、仿真结果与实际执行结果。

Performance & Robustness

技术系统课 · T09—T10

阶段产出

能够把 AI 产品部署为可持续运行的系统：支持并发模型调用、按需调度 CPU / GPU 与数据库，并以实测数据优化性能、调用成本和运行可靠性。

- 部署 SGLang + Qwen 服务，提交优化前后的吞吐、延迟与显存压测报告。
- 构建「Token 工厂」，统一管理供应商路由、密钥、配额与调用预算。
- 打通自动部署、健康监控、备份与回滚，演示故障恢复和资源回收。

T09 · 模型管理工程与 Token 效率

本节内容

- 如何构建稳定、高效、经济的 Token Server，实现调用请求的自动分配与供应商自动选择。
- 比较商用 API（OpenRouter、Bedrock 等云服务）与开源模型自部署的质量、性能和成本，以 GLM、DeepSeek 等模型为例。
- 在服务端如何为客户高效管理供应商 Token、密钥、配额、模型版本与调用预算。
- 减少无关上下文与重复调用，合理使用缓存、检索、任务拆分和模型路由；在同一验收标准下比较成本、质量与延迟。
- 理解批处理、KV Cache、并发、吞吐与延迟，以及 Flash Attention、RadixAttention。
- vLLM、Ollama 和 SGLang 的模型部署与性能优化。

课后实践案例

基于 SGLang 部署 Qwen 本地大模型服务，支持高并发调用与本地 Agent 运行；完成 KV Cache、FlashAttention 与 RadixAttention 相关配置和性能调优。在相同硬件、模型与请求负载下进行优化前后对照压测，记录并发量、吞吐、响应延迟与显存占用，提交部署代码、配置与实测报告。Qwen 的具体版本、参数规模与硬件配置另行明确。

T10 · Token 工厂：构建可持续运行的模型服务系统

本节内容

- 基于 Daytona、Render、DigitalOcean、Vercel、Supabase 和 Cloudflare 的现代化运维部署体系。
- 通过 AI 自动化管理测试、部署、密钥管理与数据库迁移等工作。
- AI 如何进行健康监控、健康检查与数据备份。
- 如何实现版本回滚与自动化部署，并结合 Git、GitHub Actions、Supabase Realtime、Vercel Cron、Render Service 等示例，理解版本管理、CI/CD、实时数据、定时任务和服务运行的分工。
- 通过 Lambda API 自动管理 GPU 资源。

课后实践案例

构建并部署「Token 工厂」：将 T09 的模型服务接入统一调用入口，管理供应商路由、密钥、配额与调用预算。按任务需求调度 CPU、GPU 与数据库资源，打通自动部署、健康检查、故障恢复与资源回收。在预算与资源供应范围内按需扩缩容，记录调用量、吞吐、延迟与成本；提交可运行的 LLM-Server 系统、部署配置、压测报告与端到端演示。

问题识别与系统接入

产品系统落地课 · B01—B02

阶段产出

理解行业与使用者，梳理已有系统、数据和流程；确定 AI 接入位置、第一版范围、验收标准及风险边界。

以真实产品与客户交付经历为驱动，拆解实际案例。

把需求、数据、接口、使用者与维护责任纳入方案；与对应阶段的技术系统课结合，完成可验收的项目产出。

B01 · AI 如何重塑经济与各个行业：什么值得做，如何避免在流沙上建房子

本节内容

- AI 在各行业发展到哪一步？覆盖医疗 AI、脑机接口、AI4Science（药物、蛋白质、材料、气象）、机器人与具身智能、自动驾驶、工业制造与能源。
- AI 在日常业务中的应用全景：金融、教育、法律、电商、内容创作、软件研发与企业运营；比较它们解决的问题、实际使用者与落地条件。
- 如何判断发展阶段？按具体场景区分研究验证、临床或业务试点、有限部署与规模化应用；以脑机接口试验、获准上市的医疗 AI 设备、AlphaFold 科研工具为例，比较能力证据、数据、监管、成本与可靠性门槛。
- AI 如何改变生产成本、竞争、定价与价值分配？
- 理解 AI 的“锯齿状技术前沿”：区分演示能力与可靠交付。
- 模型进步压力测试：如果模型原生支持核心功能，产品还剩什么？
- 识别可积累的价值：客户理解、业务集成、数据、渠道与服务能力。
- 对比案例：通用 AI 文案生成器与特定商家的内容运营系统。

背景引言

AI 正在改变各行业的生产成本、工作流程与价值分配，但不同场景的成熟度和落地门槛并不相同。结合真实应用与 Stanford、MIT Sloan、Harvard Business School 的研究，判断哪些机会值得投入，哪些产品可能是在“流沙上建房子”。

课后实践案例

选择一个行业场景，完成一页《AI 项目价值与替代风险分析》：标明发展阶段、证据来源、使用者与实际需求；判断模型进步后哪些价值仍可积累；列出三个关键假设及验证方法。商业项目进一步分析客户与付费理由。

B02 · AI 在已有系统落地接入、AI 解决真实问题：挑战与问题分析

本节内容

- 先理解现有流程：谁提出需求、谁操作系统、谁承担结果？区分表面诉求与真正的业务问题，建立当前效果、耗时与错误率的基线。
- 已有系统如何接入？盘点 API、数据库、文档、设备与人工环节，识别数据孤岛、历史系统、接口限制和信息不完整的问题。
- 数据与权限有哪些阻力？分析数据质量、更新频率、身份认证、访问范围与敏感信息边界，明确哪些数据与操作可以交给 AI。
- 模型能力如何转化为业务动作？划分确定性程序、AI 判断与人工审批，处理不确定输出、工具失败、重复执行及异常补偿。
- 怎样验证确实解决了问题？把目标写成验收指标和测试集，以只读接入、影子运行、小范围试点逐步推进，并设计回退路径。
- 如何协调交付？与业务、IT、运维和使用者确认范围、责任与依赖；以 KeploreAI 工业视觉项目分析模型、工具、Runtime 与设备接口。

背景引言

能跑通一次模型调用，并不代表接入了实际工作。本节从客户或业务现场出发，分析已有系统的约束，确定 AI 应该接入哪里、如何参与任务、怎样验收。

课后实践案例

选择一个已有系统，提交《AI 接入与问题分析方案》：画出现状流程与目标流程，列出数据、接口、权限和负责人；定义一个最小试点、验收指标、风险及回退方式。无真实系统访问条件时使用模拟接口，并注明尚待验证的假设。

试点验证与使用推广

产品系统落地课 · B03—B04

阶段产出

完成效果、成本与延迟的实测；组织小范围试点，收集使用反馈，明确预算、采用障碍与推广条件。

以真实产品与客户交付经历为驱动，拆解实际案例。

把需求、数据、接口、使用者与维护责任纳入方案；与对应阶段的技术系统课结合，完成可验收的项目产出。

B03 · AI 系统的成本、效果与交付边界

本节内容

- 拆分模型调用、CPU / GPU、数据库、人工复核与维护成本，按一次成功完成的任务核算，而不只比较 Token 单价。
- 在相同验收标准下比较质量、响应时间与成本，分析轻度、典型和重度使用的负载与预算。
- 哪些步骤值得使用 AI？选择模型路由、缓存、批处理、规则程序与人工处理，避免重复调用和无效自动化。
- 如何约定服务边界？明确支持的任务、输出质量、修改次数、超时与失败处理，以及超预算时的行为。
- 企业或内部项目如何验证收益？比较节省的工时、错误率、处理能力与持续维护成本，区分预期收益和实测结果。
- 商业化项目如何收费？比较按用量、任务、席位、订阅加额度与验收结果收费，用小范围试用验证价格与交付成本。

背景引言

系统落地需要持续承担计算和维护成本。本节围绕任务结果，决定哪些能力值得投入、服务边界如何定义，以及效果和成本如何一起验证。

课后实践案例

为自己的项目建立“效果—成本—延迟”对照表，实测三种任务负载，给出预算上限、验收条件和降级策略；商业项目补充价格与单位交付成本。

B04 · 让 AI 系统被真正使用：试点、反馈与推广

本节内容

- 确定首批使用者与业务负责人：谁有持续发生的问题，谁能够提供反馈，谁能推动系统进入日常流程?
- 如何设计小范围试点？选择真实任务，缩短首次获得价值的路径，安排培训、反馈入口与人工接管。
- 为什么上线后没人用？分析流程变化、使用成本、信任、激励与责任归属，避免只有演示、没有日常使用。
- 如何收集有效反馈？结合访谈、操作记录、失败案例与完成率，区分功能缺失、可靠性问题和不成立的需求。
- 如何从试点扩大应用？确定进入下一阶段的标准，逐步增加团队、任务与负载，并保留回退和支持机制。
- 公开产品如何找到用户？按需要使用演示、社区、内容、SEO / GEO 与用户推荐，记录“了解—试用—反馈—再次使用”的过程。

背景引言

上线只是开始。产品系统落地的结果，是使用者愿意把实际任务交给它，并在遇到问题时知道如何反馈、接管和继续工作。

课后实践案例

完成一个小范围试点计划与一次反馈复盘：写清使用者、任务、操作路径、验收指标和责任人；记录采用障碍、失败原因与下一轮改进。

行业落地与持续交付

产品系统落地课 · B05—B06

阶段产出

形成交付与持续运行手册，针对行业约束设计数据、权限、可靠性与人工接管方案；明确维护责任、预算与扩展条件。

以真实产品与客户交付经历为驱动，拆解实际案例。

把需求、数据、接口、使用者与维护责任纳入方案；与对应阶段的技术系统课结合，完成可验收的项目产出。

B05 · AI 系统的持续交付、维护与规模化

本节内容

- 如何定义交付完成？明确验收证据、数据与代码归属、部署配置、操作文档、使用培训和维护责任。
- 怎样把试点变成持续运行的服务？约定监控指标、问题分级、响应流程、模型或依赖更新与变更验收。
- 如何管理交付成本？核算计算、人工支持与维护负担，判断任务量和客户数增长时，成本与质量如何变化。
- 定制需求如何复用？区分通用能力、行业适配和客户配置，管理版本与接口，避免每接一个项目就复制一套难以维护的系统。
- 何时需要扩大团队？根据任务依赖、服务负荷和能力缺口，明确研发、产品、现场交付与客户团队的分工。
- 商业项目如何持续？结合预算、收入、回款与维护承诺制定投入计划；内部项目则明确预算来源与长期负责人。

背景引言

一个被真实场景检验的系统，需要有人负责交付之后的运行。本节把验收、维护、协作与成本组织起来，让成果能够长期使用和持续改进。

课后实践案例

为项目编制《交付与持续运行手册》：包含验收清单、部署与回滚、监控与响应、维护责任、成本预算和扩展条件，演示一次异常处理及恢复。

B06 · AI 应用在不同产业中遇到的典型问题及应对方案

本节内容

- 工业制造与机器视觉：缺陷样本稀缺、误检漏检、现场光照变化与设备协议不一致；通过数据分层、独立测试、现场校准与接口适配逐步验证。
- 企业运营、零售与物流：系统割裂、业务口径不一致、重复执行与责任不清；采用流程梳理、权限隔离、幂等操作、人工审批与审计。
- 科研与 AI4Science：数据来源、实验可复现性、计算预算与结论可靠性；保留实验配置、对照组、执行记录与独立验证。
- 医疗、金融与法律：错误代价高、信息敏感、专业责任明确；先界定使用边界，安排专业复核、权限控制与追溯，按场景确认适用要求。
- 教育、内容与 ToC 产品：质量波动、个性化偏差、版权与用户留存；建立样例标准、内容审核、来源记录与真实使用反馈。
- 如何迁移方案？比较各行业的错误成本、数据条件、实时性与接管要求，识别可复用的系统结构和必须重新验证的行业假设。

背景引言

同一个 Agent 架构，进入不同产业会遇到不同的失败方式。本节以问题和约束为中心分析案例，练习选择应对方案，而不是简单复制应用功能。

课后实践案例

任选两个产业，对比“实际问题—约束—失败模式—应对方案—验收证据”，再为自己的项目制定行业适配清单。以 KeploreAI 或 OpenLabs 的工程环节为参照，不把示例直接当成客户验收结论。

免费选修课程 · 1/2

超过 50 门课程，以下精选 20 门；配套源代码实验、计算环境与持续答疑。

01 · LLM Agents：从推理到智能体系统

来源：参考 Berkeley LLM Agents 公开课框架

02 · Agent 第一性原理：循环、上下文与 harness

来源：参考 Thorsten Ball、OpenAI Codex agent loop 与 Anthropic 上下文工程文档

03 · Agentic 设计模式（一）：提示链、路由与并行

来源：参考 Antonio Gulli《Agentic Design Patterns》第 1–3 章

04 · Agent 工程：循环、状态图与自我改进

来源：参考 Anthropic、12-Factor Agents 与 LangChain Academy

05 · 给 Agent 装能力：工具、Skill 与 MCP

来源：参考 Hugging Face MCP Course、MCP 规范与 Anthropic Agent Skills

06 · 评测 LLM Agent：pass@k、模型评委与可靠性

来源：参考 Anthropic 评测文档、τ-bench 与 LLM Evals FAQ

07 · 从零构建大语言模型

来源：参考 Stanford CS336 课程框架

08 · 从数词频到大模型

来源：OpenLabs 自编学习路线与实验

09 · 精读《深度学习》：网络怎样一层层学出有用的表示

来源：独立解读 Goodfellow、Bengio、Courville《Deep Learning》

10 · 手写自动微分引擎：造你的迷你 PyTorch

来源：OpenLabs Academy 自动微分课程；参考《Deep Learning from Scratch》等

OpenLabs 整理或独立开发；高校名称为参考框架。所列 Canvas 课程已开放首章实验，后续章节持续更新。

免费选修课程 · 2/2

超过 50 门课程，以下精选 20 门；配套源代码实验、计算环境与持续答疑。

11 · 扩散与流匹配：生成模型从零

来源：OpenLabs Academy 生成模型课程整理

12 · 生产环境的 LLM 推理服务：指标、成本与扩缩容

来源：参考 BentoML / Modular 推理手册、Google Cloud GKE 与《Inference Engineering》

13 · 精读《AI 系统性能工程》：让 GPU 的每一秒都干有用的活

来源：独立解读 Chris Fregly《AI Systems Performance Engineering》

14 · GPU 并行编程：从 SIMT 到写出快的 kernel

来源：参考 Kirk & Hwu《Programming Massively Parallel Processors》等

15 · 机器学习工程化：从模型到生产系统

来源：参考 CMU ML in Production 课程框架

16 · 具身智能：让程序长出身体，在世界里把事做成

来源：OpenLabs Academy 具身智能课程与模拟实验

17 · 现代机器人学：让机器人动起来的几何

来源：OpenLabs Academy 现代机器人学课程整理

18 · 自动驾驶入门：从循迹到 mini 规划栈

来源：OpenLabs Academy 自动驾驶课程整理

19 · 深度强化学习：从 MDP 到策略梯度与探索

来源：参考 Stanford CS234 强化学习课程框架

20 · 计算机视觉：从像素到卷积

来源：OpenLabs Academy 视觉课程；参考 Szeliski、Forsyth & Ponce 相关教材

OpenLabs 整理或独立开发；高校名称为参考框架。所列 Canvas 课程已开放首章实验，后续章节持续更新。

往期学员项目 · 内容与办公自动化

Cohort 1 / 2 · 项目成果与技术栈分析

自动录制与剪辑视频助理

把主题变成口播稿，控制 OBS 录制，再完成语义剪辑和字幕。

主题与资料 → 口播稿 / 提词器

OBS 录制 → 转录与选段

剪辑 → 字幕 → 视频成片

报告流程示意 · 含真人录制与剪辑记录

技术栈与设计

Python Agent Loop · OBS WebSocket · FFmpeg

模型选择文本片段，工具读取真实时间戳；超额删减先确认，减少误剪。

实验报告记录：38.55 秒素材 → 7.217 秒剪辑片段。

飞书周报与月报助理

拉取工作数据，自动整理部门、团队周报，并按日期触发月报。

飞书工作数据

部门周报 / 团队周报 / 月报

定时检查 → 已生成则不重复

功能结构示意 · 报告记载实际生成成功

技术栈与设计

Feishu Open API · Skills · Cron / Heartbeat

把日期规则、历史状态和格式模板分开，减少重复提醒与手工整理。

学员复盘：8 月 6 日实际生成；省去手动整理。

个性化行业资讯播报

把通用播报改成个人关注的行业主题，持续推送到邮箱或飞书。

个人主题 → 资料抓取

筛选、摘要与来源链接

邮箱 / 飞书 → 定时送达

多位学员的实践方向汇总

技术栈与设计

MCP 检索工具 · LLM 摘要 · 邮件 / 飞书 · Cron

把内容来源、摘要与推送拆开；数据源受限时调整采集方式。

有个人开发日志；学员直接回复确认收到推送。

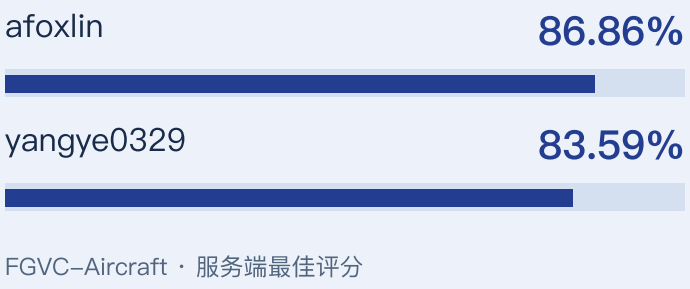
依据往期实验报告与评测记录整理（2026.10）；流程图为成果结构示意，不是产品截图。

往期学员项目 · 模型训练与软件研发

Cohort 1 / 2 · 项目成果与技术栈分析

能够自动训练微调 GPU 视觉模型的 Agent 系统

让 Agent 驱动 GPU 视觉模型的训练与微调，读取评测结果，比较和调整训练策略。



FGVC-Aircraft · 服务端最佳评分

技术栈与设计

Agent 工具调用 · GPU 训练 · 微调 / TTA

用评测反馈比较训练策略。两位学员分别留下 11 次、8 次挑战提交记录。

分数来自平台评分；优化方法以各自实验报告为准。

mini Claude Code

自己构建 Coding Agent，让它读取代码、调用工具、修改并测试修复。

- 任务 → 上下文与工具调用
- 修改代码 → 执行测试
- 本地课程修复任务：3 / 3

本地 runner 上传成绩 · 非完整 SWE-bench

技术栈与设计

Agent Loop · 工具路由 · 权限策略 · 子 Agent

模型生成能力之外，还要管理上下文、工具分发与执行权限。

有本地成绩上传与实验报告；服务器未独立重跑。

WPS 文档自动化助理

给常驻助理增加办公能力，生成文档与演示内容，并导出 PDF。

- 任务请求 → 本机助理
- Windows WPS COM 自动化
- 文档 / 演示内容 → PDF

扩展任务示意 · 来源于学员实验报告

技术栈与设计

Windows · WPS COM · 本机工具 · 守护任务

区分平台接口、约束输出目录，并增加重启与恢复机制。

报告包含 WPS 接口适配和自动化扩展过程。

依据往期实验报告与评测记录整理（2026.10）；流程图为成果结构示意，不是产品截图。

往期学员项目 · 业务系统与可靠性

Cohort 1 / 2 · 项目成果与技术栈分析

常驻邮件与 Slack 助理

筛选重要邮件，推送给用户；在 Slack 修改回复草稿并人工确认发送。

邮件监听 → 重要性判断

Slack 草稿 → 人工确认

发送 / 状态保存 / 断线恢复

流程示意 · 外部动作保留人工确认

技术栈与设计

IMAP IDLE · Slack · 状态文件 · 任务守护

用处理位置和状态文件避免重复消费；守护进程负责持续运行。

多份实验报告，以及学员对邮件筛选的具体反馈。

企业零售客服 Agent

连接业务环境与 GitHub CI，用业务测试检查 Agent 的实际行为。

集成任务	1 / 1 通过
另一业务套件	10 / 11

rivalzhao 的不同评测记录；业务套件整体未通过

技术栈与设计

业务 API · GitHub Actions · 自动化业务评测

连通性验证与业务验收分开；测试暴露硬编码和不符合业务约束的实现。

有仓库、CI 与平台评测回执；展示迭代过程。

可恢复与自我改进 Agent

让任务能停、能续、能人工审批，并用独立评测检查每次修改。

执行 → 检查点 → 中断恢复

候选修改 → 独立 Judge

通过才保留；失败不晋级

机制示意 · 学员报告也记录了无改进与失败

技术栈与设计

检查点 · 人工审批 · 独立 Judge · 留出测试

将执行状态与评分边界隔离，避免重复调用，也避免 Agent 修改记分规则。

有恢复实验与自改反思，不把每轮修改视为提升。

依据往期实验报告与评测记录整理（2026.10）；流程图为成果结构示意，不是产品截图。

往期学员反馈

“我觉得，高老师的每一堂课都是精品。讲的都是自己实战中获得的经验，深入浅出，理论结合实战做得很好。”

—— Z同学

“我也上过其他相关课程，感觉偏科普，讲得比较浅。还是高老师专业。”

—— Y同学

“我看到 Manus 2.0 时，感觉就是高老师教我们的内容——一个 Agent 综合体。”

—— 周同学

“用了自动化测试，我才发现之前 Agent 写的代码里很多地方都硬编码了，也发现了之前实现中的问题。”

—— 樱同学

根据学员原话轻微整理，保留第一人称语气；樱同学的测试体会结合导师补充说明。

资源价值账单：把项目做起来需要哪些支持

资源与支持	本计划提供	独立采购 / 核算参考
GPU 算力	每人最多 100 GPU 小时；机型按任务分配	示例：100 小时 H100 SXM × US$4.29/小时 = US$429
模型 Token	每位学员 US$50 Token 额度	US$50；按模型和实际调用用量消耗
项目与学习材料	9 项课程练习 + 3 项专题实战（配环境与指导）	选修免费开放，配套源码、环境与答疑
产品与研发指导	16 次直播；答疑与个人项目指导按方案安排	按服务周期与约定交流安排确认
半年部署与陪跑	项目落地与半年陪跑方案提供 CPU、数据库、GPU 自动运维部署与 Stripe 接口，持续 6 个月	课程方估算 US$100–200/月 × 6 个月 = US$600–1,200

半年陪跑方案 · 示例资源预算 US$1,079–1,679

另含 16 次直播、一对一指导与 9 个项目源码；50+ 门选修课程免费开放。

参考：Lambda H100 SXM 单卡价格（2026-10-20，未含税）。合计按 100 小时全部使用 H100 的假设测算；实际机型与用量不同，不能视为现金返还或保证节省金额。合计含半年服务，仅对应 US$1,880 方案；运维估算含 CPU、数据库及服务，不重复计 GPU 与 Token。

Fullstack AI Engineer & FDE Training Program

训练营正式开始时间：2026年11月14日

导师：高老师（Marvin Gao） · OpenLabs / Agentist

费用与支持	课程与实战方案	项目集训方案	项目落地与半年陪跑方案
项目学费	**US$735** 约合人民币 ¥4,800	**US$1,180** 约合人民币 ¥7,800	**US$1,880** 约合人民币 ¥12,800
共同学习支持	完整直播课程 · 完整项目练习 · 计算资源支持 自动化评测系统使用 · 20 多门精品课程 本期三项实战 · Enterprise AI Agent / LLM-Server / Embodied AI（配环境与指导）		
1 对 1 支持	课程与项目练习答疑	课程期间 一对一个人项目指导	一对一定制项目指导 与落地陪跑
定制项目陪跑	**不含定制项目陪跑**	课程期间指导个人项目 不含半年持续陪跑	**含 6 个月** **定制项目陪跑**
半年部署服务 与支付接入	不含此项半年服务	不含此项半年服务	服务器、数据库、GPU 自动运维部署及 Stripe 接口
报名退费	**报名后 15 天内，可无理由全额退费。**（三种报名方案均适用）		
项目上线保障	不含上线专项保障	不含上线专项保障	未达约定上线目标，按约定条件 退还 **¥5,000 或 US$700** 学费

以美元计价，人民币为参考金额，随当日汇率变化。上线期限与验收标准、半年服务起止时间与资源范围、交流安排及退费流程，在报名确认时书面约定。

扫码填写申请表

先申请，与老师一对一交流

交流完毕、确认方案后再付费

Apply

为保证质量，此次仅招收

15名学习者参与