

工信部教考中心《大模型开发工程师》简介

一、项目简介

大模型开发工程师旨在精准评估和考核在大模型开发领域具备专业技能与实践经验的人才，聚焦于大模型架构设计、算法研发、模型训练与优化、部署及创新应用等核心能力，全面覆盖从基础理论到前沿技术实践的各个层面，有力推动人工智能技术在各行业的深度应用与创新发展。

证书分为初级、中级、高级三个等级，为不同阶段的从业者提供科学、系统的能力评价体系。

二、能力标准

1、大模型开发工程师（初级）

了解大模型的基本概念、发展脉络和主流架构，熟悉常见的深度学习框架的基础操作。

能够运用框架完成简单模型的搭建与训练，掌握基本的数据处理技能，确保数据能够适配模型训练需求。

具备一定的编程基础，编写模型训练和数据处理的基础脚本，能够完成小型模型开发项目中的简单模块任务。

2、大模型开发工程师（中级）

熟练掌握深度学习框架的核心功能和高级特性，能够独立设计和开发中等复杂度的模型架构。

深入理解模型训练的原理和优化策略，能够运用多种技术手段有效提升模型性能。具备较强的数据处理和分析能力，可进行复杂数据的预处理、特征工程和数据分析，为模型开发提供高质量的数据支持。

3、大模型开发工程师（高级）

精通大模型开发的全流程技术，包括先进架构设计、高效训练算法研发和精准的模型微调技巧。

熟练运用多种深度学习框架和工具，深入了解微调工具等专业开发平台的高级功能，能够根据不同的应用场景和硬件环境进行高效的模型开发和优化。

在大规模模型开发项目中，能够发挥核心技术引领作用，带领团队攻克技术难题。具备深厚的数学和算法基础，能够对模型性能进行深入分析和优化，提出创新性的解决方案。

三、课程内容

阶段	课程	知识模块
基础知识	大模型与AIGC概述	1. 大模型是什么 2. 大模型的前世今生 3. 大模型爆火的原因 4. 大模型原理介绍 5. AIGC 简介
	PyTorch框架基础实践	1. PyTorch 简介 2. 张量操作 3. 构建一个线性模型 4. 识别手写数字
LangChain 与大模型应用开发	LangChain实战	1. LangChain 介绍与安装 2. LangChain 快速使用 3. 模型 IO 过程解析 4. RAG 实践：开发一个课程助教机器人 5. 代理（Agents）

	基于LangChain的本地大模型调用	1. HuggingFacePipeline 使用 2. 自定义 LLM 3. 函数式链的使用 4. 基于 Embedding 的文本相似度计算 5. 在线大模型使用 6. 情感分析应用实战 7. 命名实体识别实现
	RAG实战-论文阅读助手	1. 背景与目标 2. 模型准备 3. 文档读取和预处理 4. 向量存储 5. 构建阅读问答链 6. RAG 增强技术 7. 阅读机器人界面开发
	Agent实战-在线教育课程订单智能助手	1. 背景与目标 2. 模型准备 3. 创建表格智能体 4. Agent 创建和使用
	基于LangChain的人事面试助手开发	1. 目标分析 2. RAG 技术流程 3. LangChain 加载本地大模型 4. 文档读取和分割实现 5. 构建 RetrievalQA 问答链 6. Gradio 框架
大模型部署与微调	大模型开源平台介绍及使用	1. LLM 大模型基础认知 2. 大模型部署与应用技术概述 3. HuggingFace 开源 AI 社区 4. 小结
	PyTorch大模型原理与应用	1. 基础模型简介 2. 万模基座 Transformer 3. 生成式预训练模型 GPT 4. 双向编码模型 BERT 5. Transformer 应用
	大模型微调原理与实现	1. 大模型微调的概念 2. 大模型微调方法 3. 全量微调 FFT 4. 参数高效微调 PEFT 5. 如何降低大模型显存占用 6. modelscope 平台 7. LLaMA-Factory 界面微调工具
	基于大模型的金融问句语义相似度计算	1. 案例背景及挖掘目标 2. 语义文本相似度

		3. 预处理 4. 加载预训练模型 5. 定义损失函数 6. 模型训练 7. 模型预测 8. 模型评估 9. 小结
	基于大模型微调的命名实体识别	1. 背景与目标 2. 命名实体识别概述 3. 模型准备 4. 数据准备 5. 微调训练和验证 6. 小结
	基于通义千问(Qwen)的创意广告生成	1. 背景与目标 2. 千问大模型 3. 数据准备 4. 模型微调和验证 5. 小结
	基于BERT模型的中医问答系统	1. 案例背景与挖掘流程 2. 数据加载 3. 数据探索：三元组数据长度分布情况 4. 预训练模型介绍 5. 模型加载 6. 加载本地 <code>checkpoint</code> 文件并进行预测 7. 小结
	部署自己的Qwen2大模型	1. Github 代码托管平台 2. Qwen2 大模型部署 3. Qwen2 大模型量化 4. 其他大模型部署示例
	Qwen2大模型部署与电商客服模型微调	1. 背景与目标 2. 环境准备 3. 数据格式转化 4. 执行微调脚本 5. 微调后的模型使用效果分析 6. web 界面部署 7. swift 模型推理 8. swift 界面化微调训练
	基于Ollama和Dify的本地大模型的部署和应用开发	1. 背景与目标分析 2. Ollama 介绍与安装 3. 通过 openai 调用 ollama 本地大模型 4. Ollama 部署和使用 Embedding 模型 5. 通过 ModelFile 部署本地大模型

		6. LM Studio 介绍与安装 7. LM Studio 使用 8. Dify 平台介绍与安装 9. 基于 Dify 的 RAG 实现流程 10. 创建通用聊天机器人
职业技术考核	大模型开发工程师（高级）职业技术考试	

四、报考条件

初级：无要求，皆可报考。

中级：（满足其中一个条件即可）

- 1.获得初级证书。
- 2.年满18周岁，具备高中以上学历，工作年限满1年。

高级：（满足其中一个条件即可）

- 1.获得中级证书。
- 2.年满20周岁以上，工作年限满2年。
- 3.年满20周岁以上，具备大专及以上学历。

五、考试方式

考试方式分为线下考试站点或线上考试系统统考两种方式，考试形式为上机答题，闭卷。

考试题型：

初级：120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

中级：120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

高级：120 分钟，单选题+多选题+判断题+操作题+简答题，上机答题。

六、成绩评分

考试最终成绩满分为100分，成绩80-100分为优秀；成绩60-79分为合格；成绩60分以下为不合格。

七、证书样本

学员经考核合格，由工业和信息化部教育与考试中心颁发大模型开发工程师
职业技术证书，证书可登录国家工信部教育与考试中心官网查询。

