

1 AI 及 LLM 基础

- 1. AI 领域基础概念
 - 学习目标: 建立对AI和LLM的基础理解, 了解主要的机器学习和神经网络模型, 掌握API调用方法。
 - AI, NL/NLU/NLG
机器学习: 学习方法, 拟合评估
神经网络: CNN, RNN, Transformer
Transformer, BERT 架构简介
LLM, Generative AI
OpenAI 及其他常用大模型介绍演示
参数详解: max_tokens, temperature, n, top_p, presence_penalty, frequency_penalty
API Key | OpenAI API 调用
- 2. OpenAI 开发
 - 学习目标: 学会使用OpenAI API开发聊天机器人, 理解API的参数和计费, 掌握不同API调用方法。
 - OpenAI API 参数: Token, 计费
Rest API 调用(Postman): Model, Chat, Audio, Images, Embedding
Python API 调用
Tool calls
调用具有视觉的 GPT-4o
使用本地图片
JSON 模式
seed 可重现输出
使用代码统计token 数量
开发控制台循环聊天
管理对话 token: 实现基于最大 token 数量的消息列表限制
带会话长度管理的控制台循环聊天
带会话长度管理的控制台 Chatbot 示例
- 3. 支持多模态输入的 AI Chatbot App
 - 学习目标: 开发支持图片等多模态输入的Chatbot应用, 优化代码结构以提高开发效率。
 - 开发基于 Streamlit 和 OpenAI 大模型的 Chatbot App
支持图片的多模态输入
Python 项目代码结构优化
基于 Streamlit 的 Web UI 多模态 Chatbot
- 4. Prompt Engineering 提示词工程
 - 学习目标: 学习如何设计高效的Prompt以获得最佳模型输出结果, 深入掌握Prompt Engineering技巧, 提升与大模型的交互质量。
 - 从事岗位: 提示词工程师 (15K-30K)
 - Prompt engineering 介绍
最佳实践
Zero-shot, Few-shot
CoT, ReAct, Reflexion
Prompt Chaining
结构化输出
Prompt 应用: 数据, 代码, 函数, 分类, 创意, 生成, 汇总
风险与安全
gpt4-version 图像提示工程技术

2 LangChain 开发

- 学习目标: 掌握LangChain的高级工具与RAG和Agent技术, 能够构建复杂的AI应用。
- 从事岗位: AI大模型应用开发工程师 (30K+)
- 1. LangChain 基础
 - 学习目标: 掌握LangChain的基本用法, 理解如何在复杂任务中使用框架。
 - LangChain 模块和体系
LLM & Chat models
Prompt(Templates, OutputParsers)
Chains
LCEL & Runnable interface
LLM apps debug: Verbose, Debug Mode & LangSmith Tracing
- 2. LangChain Chat Model
 - 学习目标: 能够使用LangChain开发具有记忆功能的Chatbot。
 - Prompt templates: Few shot
LangServe
为 Chain 添加 Message history (Memory)
单行初始化 chat model
基于 LangChain 的 Chatbot: Chat History
Track token usage
Cache model responses
- 3. LangChain Tools & Agent
 - 学习目标: 深入学习LangChain的工具和Agent, 包括数据处理、任务管理、插件使用等。
 - Multimode 集成
Output parsers: JSON, XML, YAML
自定义 Tools, 调用 Tools
集成内建 Tools
创建和运行 Agent
- 4. Embedding & Vector Store
 - 学习目标: 掌握如何使用Embedding模型和向量数据库进行高效的信息检索。
 - Embedding models
Vector stores: 向量数据库: Chroma, Weaviate, Qdrant, Milvus, Pinecone, FAISS, Lance
Similarity search
向量检索: Document, Retrievers
基于 LangChain 构建向量存储和查询
开发基于 Embedding 的意图匹配和问答实例
- 5. RAG 专题
 - 学习目标: 理解RAG工作机制, 学会使用文档加载器和文本分割器处理数据, 学习如何结合检索和生成模型来处理复杂的任务。
 - RAG 简介
Document loaders: CSV, HTML, JSON, Markdown, PDF
Text splitters: recursively, HTML, code, Markdown, JSON, semantic chunks, tokens
Loader, Splitter, Embedding, Vector Store, Retrievers 的综合应用
- 6. 自定义组件专题
 - 学习目标: 掌握回调处理和自定义模型的开发。学会自定义组件以满足特定需求, 增强模型的灵活性和功能性。
 - Callback 处理
自定义 callback handlers
自定义 Chat model
自定义 RAG: Retriever, document loader
- 7. 基于 LangChain 的常用案例实战
 - 学习目标: 掌握使用LangChain处理实际问题的方法, 包括文档问答、数据查询和集成网络搜索功能。
 - PDF 文档问答
基于 Web URL 的问答
基于 SQL 的 CSV 数据分析问答
Web Search 集成
- 8. 带 RAG 的 Chatbot 实战
 - 学习目标: 构建一个具有复杂对话记录处理能力的Chatbot。
 - Chatbot Memory
RAG
Tools
大对话记录处理
带 RAG 的 Chatbot
- 9. LangGraph 开发实战
 - 学习目标: 掌握LangGraph的应用, 理解如何通过图结构来增强模型的推理能力, 学会构建复杂的多智能体系统。
 - LangGraph 核心组件: Graphs, State, Nodes, Edges, Send...
LangGraph 实现: 可控性 Agent 框架, 持久化, Human-in-the-loop, streaming, React agent
Agent 使用案例: Chatbots, Multi-Agent Systems, RAG, Planning Agent...

3 LlamaIndex 开发

- 1. Llamaindex 框架
 - 学习目标: 掌握LlamaIndex的核心模块和应用构建。
 - 与 LangChain 框架对比
核心模块介绍: Models, Prompts, Loading, Indexing, Storing, Querying, Agent, Evaluation
RAG pipeline 构建: Loading, Indexing, Querying, Evaluating
Agent 构建
- 2. 基于 Llamaindex 框架的应用
 - 学习目标: 基于LlamaIndex框架开发大规模文本搜索和内容推荐系统。
 - 大规模文本数据搜索
内容推荐
AI 问答
LlamaIndex 与 LangChain 集成
LlamaIndex RAG CLI 工具

4 AutoGen 开发

- 1. AutoGen Studio
 - 学习目标: 学习使用AutoGen studio进行工作流设计与管理, 理解其核心组件及功能。
 - AutoGen Studio UI
AutoGen Studio workflow
Python 集成 workflow
- 2. AutoGen Framework
 - 学习目标: 掌握AI代理的开发与集成, 如何通过代理来增强模型的应用场景。使用AutoGen进行自动化代码生成和多智能体对话。
 - 快速入门
LLM 配置
Agent 类型: UserProxy, Assistant, Conversable Agent
Code Executors 自动化代码
Tool Use
GroupChat 多 Agent 对话

5 LLM 大模型训练与微调

- 学习目标: 具备大模型的训练与调优能力, 能够开发和优化适用于特定领域的AI模型。
- 从事岗位: AI大模型工程师 (40K+)
- 1. 模型即服务共享平台
 - 学习目标: 学会使用共享平台和本地环境开发和部署大模型语言模型。
 - Hugging Face, ModelScope 平台功能
加载在线模型, 离线模型
模型推理调用
Transformers 使用
利用模型完成任务: 文本, 语音, Embedding Models
- 2. 私有部署本地大模型
 - 学习目标: 掌握本地大模型的安装部署和调用。
 - 硬件配置分析 (推理和微调)
常用大模型选型: LLaMA 3, Gemma, Mistral...
本地大模型安装部署
本地大模型调用
本地大模型与对话 APP 应用集成
- 3. 训练自己的大模型
 - 学习目标: 深入理解Transformer模型的工作原理, 掌握大模型的训练流程和评估, 量化技术。
 - Transformer 模型详解, Encoder, Decoder
制作和加载数据集 (训练、验证、测试)
模型训练Train (单卡、多卡), 分布式训练
模型评估: 方法论、评估指标
模型量化: AWQ, GPTQ, GGUF
模型部署
API 调用
- 4. 大模型微调
 - 学习目标: 掌握基于预训练模型的微调技术, 针对特定领域进行模型的训练与优化, 构建专属行业模型。
 - 构建专属行业模型: 基于预训练大模型 Fine-tune
微调对比: 提示工程, RAG, Fine-tuning, QLoRA
Prompt Tuning, Prefix-Tuning, LoRA, QLoRA
基于 OpenAI Fine-tuning
本地大模型的 Fine-tuning
特定任务的大模型推理和训练实践: 文本分类, 序列标注, 文本生成, 图像分割
- 5. 特定任务的模型微调训练
 - 基于BERT的中文评价情感分析
如何处理超长文本训练问题
GPT2-中文生成模型定制化微调训练
GPT2-中文生成模型定制化内容输出
LLaMA3大模型本地部署与调用
使用自定义数据集和LLaMA-Factory完成LLaMA3微调训练
LLaMA3 LoRA数据测试评估、模型合并、量化
LoRA微调后的模型部署
- 6. 多模态大模型使用
 - 多模态大模型基本概念
本地部署CogVideoX-5B文生视频模型
本地 Llama-3.2-Vision模型架构剖析
Llama 3.2-Vision模型预期用途
使用ollama部署Llama-3.2-11B-Vision-InstructGGUF实现视觉问答

6 企业级项目实战

- 学习目标: 通过企业级真实项目场景落地, 锤炼大模型全栈技术应用能力。完成AI应用从策划到开发到落地的全过程。
- RAG项目实战 (基于 Llamaindex 构建企业私有知识库)
 - RAG项目背景介绍
大模型企业级业务场景落地方案实践
使用Conda配置项目环境
Sentence Transformer 模型详解
Embedding 文本向量化处理
InternLM2 1.8B / Qwen2.5 0.5B模型实战
使用HuggingFace调用本地模型
知识库模型问答测试与实际效果评估
创建Web应用
自定义知识库设置
- 基于本地大模型的在线心里问诊系统 (微调项目)
 - AI大模型项目开发流程
基座模型选择
训练数据集制作
Xtuner微调大模型
模型评估
模型量化 Weight Auto
模型部署: LMDeploy
静态和动态推理性能
推理服务 API Server
服务部署
模型转换与合并
模型打包、部署
- 基于YOLO的骨骼识别项目 (视觉项目)
 - 目标检测与图像识别概念
YOLO检测架构分析
YOLO模型的基本使用
目标检测模型数据标注与模型训练
使用自定义数据集调用YOLO的多目标检测
骨骼检测项目流程分析
voc数据集转为yolo数据集
手骨关节检测模型
骨骼分类模型
整合目标检测与图像识别模型实现骨骼识别
- 企业业务需求-政府采购采购项目 (新)
 - 企业基础数据收集
配置 Qwen2.5 3B模型
数据集转换
使用 Llama Factory 进行模型微调
open compass模型评估
测试 Qwen2.5 3B 模型微调效果
基于 Llamaindex 实现 RAG
模型训练获取模型的
构建知识库图谱“标的物”
应用场景1: 采购企业, 热门采购物资市场查询
应用场景2: 供应商, 查找采购企业
应用场景3: 技术研发、技术方向、设计材料
- 基于大模型的语音唤醒项目实战 (新)
 - 语音应用场景
PyTorch语音模块介绍
PyTorch解析音频数据
使用 Wav2Vec2 实现语音识别
- AI虚拟数字人项目正在测试阶段...
- 更多AI大模型项目每周持续增加中...

7 前沿技术扩展

- 学习目标: 了解并掌握AI领域的前沿技术, 包括语音识别、图像处理、目标检测和多模态学习。
- 1. OpenAI Assistants 开发
 - Assistant 关键组件
创建 Assistant, Thread, 消息处理
工具集成: Code interpreter, Function calling, File search, Vector store
- 2. 音频模型专题
Large Audio Language Model
 - 中英文语音识别
语音翻译
语音标题生成
带词级时间戳的语音识别
音频场景分类
音频情感识别
问答
- 3. 人脸模型专题
Face detection Model
 - 人脸检测
人脸活体
人脸关键点
人脸属性&表情, 表情识别
人脸识别, 人脸认证、通行等
- 4. 图片处理模型专题
Text to Image/Stable Diffusion Model
 - Text-to-Image Diffusion Models 扩散模型
漫生风, 水彩风, 剪影生成, 卡通系列
文生图, 图生图, 图生图, 图生图
抠图
stable video diffusion 视频生成模型
图片生成视频
画质增强, 人像修复, 图像去模糊, 去噪
- 5. 检测模型应用专题
Object Detection Model
 - 垂直类检测: 安全帽, 香烟, 口罩, 人体, 人头
视频目标检测跟踪: 自动驾驶, 交通标志, 行人
行为检测
关键点检测: 人脸, 人手, 人体
- 6. 行业信息处理专题
Structbert Model
 - 零样本学习: 抽取式, 生成式
信息抽取: 分词, 谓词标注, 命名实体识别, 关系抽取, 事件抽取, 关键词抽取, 匹配, 分类, 指代消解, 阅读理解
- 7. 多模态应用专题
MultiModal
 - 图像描述
视频描述
视觉定位
文本生成视频
多模态表征

更多AI前沿技术持续更新中...

Preview