# 《软件2.0——AI大模型企业级应用开发实战》

陈光剑 编著

AI 天才研究院 / AI Genius Institute, 2024

# 第一部分: 软件2.0时代的AI大模型基础

# 第1章: 软件2.0与AI大模型概述

## 1.1 软件 2.0 时代的到来

### 1.1.1 从软件 1.0 到软件 2.0 的演进

### 1.1.2 大模型在软件 2.0 中的核心地位

### 1.1.3 企业级应用开发的新范式

## 1.2 AI大模型的定义与特点

### 1.2.1 AI大模型的定义

### 1.2.2 AI大模型的核心特点

### 1.2.3 AI大模型与传统AI的区别

## 1.3 主流AI大模型简介

### 1.3.1 GPT系列模型

### 1.3.2 BERT及其变体

### 1.3.3 其他知名大模型介绍

## 1.4 AI大模型在企业中的应用前景

### 1.4.1 AI大模型的潜在应用领域

### 1.4.2 企业采用AI大模型的优势

### 1.4.3 AI大模型应用的挑战与机遇

# 第2章: AI大模型技术基础

## 2.1 深度学习与神经网络基础

### 2.1.1 神经网络的基本结构

### 2.1.2 常见的深度学习架构

### 2.1.3 深度学习优化算法

## 2.2 自然语言处理技术概览

### 2.2.1 词嵌入技术

### 2.2.2 序列模型与注意力机制

### 2.2.3 转换器架构详解

## 2.3 大规模预训练模型原理

### 2.3.1 预训练的概念与意义

### 2.3.2 自监督学习方法

### 2.3.3 迁移学习与微调技术

# 第二部分: 企业级AI应用开发准备

# 第3章: 企业AI战略规划

## 3.1 AI在企业中的价值评估

### 3.1.1 AI能力评估框架

### 3.1.2 AI项目ROI分析方法

### 3.1.3 AI对企业业务流程的影响

## 3.2 制定企业AI应用路线图

### 3.2.1 AI应用场景识别与优先级排序

### 3.2.2 分阶段AI实施计划

### 3.2.3 资源分配与预算规划

## 3.3 AI项目管理最佳实践

### 3.3.1 AI项目生命周期管理

### 3.3.2 跨职能团队协作模式

### 3.3.3 AI项目风险管理策略

# 第4章: AI开发环境搭建

## 4.1 硬件与软件需求分析

### 4.1.1 GPU服务器配置指南

### 4.1.2 分布式计算环境搭建

### 4.1.3 开发工作站规格建议

## 4.2 开发框架与工具选择

### 4.2.1 深度学习框架对比(TensorFlow, PyTorch等)

### 4.2.2 NLP工具包介绍(NLTK, spaCy等)

### 4.2.3 模型服务部署工具(Docker, Kubernetes等)

## 4.3 云端vs本地部署方案对比

### 4.3.1 公有云AI服务评估

### 4.3.2 私有云部署方案设计

### 4.3.3 混合云策略制定

# 第5章: 数据准备与预处理

## 5.1 企业数据资源评估

### 5.1.1 数据资产盘点方法

### 5.1.2 数据质量评估标准

### 5.1.3 数据价值分析框架

## 5.2 数据清洗与标注技术

### 5.2.1 结构化数据清洗流程

### 5.2.2 非结构化数据预处理技术

### 5.2.3 数据标注工具与方法

## 5.3 数据安全与隐私保护策略

### 5.3.1 数据脱敏技术

### 5.3.2 差分隐私在数据处理中的应用

### 5.3.3 数据访问控制与审计机制

# 第三部分: AI大模型应用开发实战

# 第6章: 智能客服系统开发案例

## 6.1 需求分析与系统设计

### 6.1.1 智能客服系统功能需求分析

### 6.1.2 系统架构设计

### 6.1.3 用户交互流程设计

## 6.2 模型选择与微调

### 6.2.1 预训练模型选择策略

### 6.2.2 领域特定数据收集与处理

### 6.2.3 模型微调技术与实践

## 6.3 对话管理与意图识别实现

### 6.3.1 意图分类模型训练

### 6.3.2 实体识别与槽位填充

### 6.3.3 对话状态跟踪实现

## 6.4 多轮对话与上下文管理

### 6.4.1 上下文表示方法

### 6.4.2 长短期记忆机制设计

### 6.4.3 多轮对话生成策略

## 6.5 系统集成与性能优化

### 6.5.1 与企业CRM系统集成

### 6.5.2 实时响应优化技术

### 6.5.3 A/B测试与持续改进方法

# 第7章: 智能文档分析系统开发案例

## 7.1 文档处理流程设计

### 7.1.1 文档类型识别

### 7.1.2 文档预处理pipeline设计

### 7.1.3 文档解析策略制定

## 7.2 OCR与文本提取技术应用

### 7.2.1 OCR引擎选择与集成

### 7.2.2 图像预处理技术

### 7.2.3 后处理与文本校正方法

## 7.3 文档分类与信息抽取模型开发

### 7.3.1 文档分类模型训练

### 7.3.2 命名实体识别在文档中的应用

### 7.3.3 关系抽取技术实现

## 7.4 知识图谱构建与应用

### 7.4.1 领域本体设计

### 7.4.2 知识抽取与图谱构建

### 7.4.3 知识推理与查询优化

## 7.5 系统评估与迭代优化

### 7.5.1 文档处理准确率评估方法

### 7.5.2 系统性能指标设计

### 7.5.3 用户反馈收集与模型更新机制

# 第8章: 智能推荐系统开发案例

## 8.1 推荐算法选择与设计

### 8.1.1 协同过滤算法实现

### 8.1.2 基于内容的推荐方法

### 8.1.3 深度学习推荐模型设计

## 8.2 用户画像构建

### 8.2.1 用户特征工程

### 8.2.2 行为序列建模

### 8.2.3 动态兴趣捕捉技术

## 8.3 个性化推荐模型训练

### 8.3.1 样本构建与负采样策略

### 8.3.2 多目标优化训练方法

### 8.3.3 冷启动问题解决方案

## 8.4 A/B测试与效果评估

### 8.4.1 离线评估指标设计

### 8.4.2 在线A/B测试方案

### 8.4.3 长期效果跟踪与分析

## 8.5 实时推荐引擎部署

### 8.5.1 特征服务设计

### 8.5.2 实时预测服务架构

### 8.5.3 缓存策略与性能优化

# 第9章: 电商导购助手项目

## 9.1 项目需求分析与系统设计

### 9.1.1 电商场景下的AI导购需求分析

### 9.1.2 导购助手系统架构设计

### 9.1.3 用户交互流程规划

## 9.2 商品知识图谱构建

### 9.2.1 商品数据采集与预处理

### 9.2.2 商品属性抽取与关系建模

### 9.2.3 知识图谱存储与查询优化

## 9.3 自然语言理解模块开发

### 9.3.1 意图识别模型训练

### 9.3.2 实体识别与属性抽取

### 9.3.3 上下文管理策略

## 9.4 个性化推荐算法实现

### 9.4.1 基于知识图谱的推荐方法

### 9.4.2 实时用户兴趣建模

### 9.4.3 多样性与新颖性平衡策略

## 9.5 对话生成与多模态交互

### 9.5.1 基于模板的回复生成

### 9.5.2 商品图文信息融合展示

### 9.5.3 交互式导购对话流程设计

# 第10章: 智能代码插件工具项目

## 10.1 开发环境与插件架构设计

### 10.1.1 目标IDE插件开发框架选择

### 10.1.2 插件功能模块设计

### 10.1.3 与AI服务的通信架构

## 10.2 代码理解与静态分析

### 10.2.1 抽象语法树(AST)解析实现

### 10.2.2 代码结构与依赖分析

### 10.2.3 代码语义理解模型训练

## 10.3 智能代码补全功能实现

### 10.3.1 上下文感知的代码补全模型

### 10.3.2 实时补全候选生成算法

### 10.3.3 用户编码习惯学习与适应

## 10.4 代码重构与优化建议

### 10.4.1 代码坏味道检测

### 10.4.2 自动重构规则设计与实现

### 10.4.3 性能优化建议生成

## 10.5 自然语言到代码转换

### 10.5.1 需求描述解析

### 10.5.2 代码生成模型fine-tuning

### 10.5.3 生成代码的质量控制与优化

# 第11章: 智能数据报表项目

## 11.1 数据报表需求分析与设计

### 11.1.1 企业报表需求调研方法

### 11.1.2 报表系统架构设计

### 11.1.3 数据可视化策略制定

## 11.2 数据处理与分析pipeline

### 11.2.1 多源数据集成与清洗

### 11.2.2 数据仓库与数据湖设计

### 11.2.3 实时数据处理流架构

## 11.3 智能数据分析模型开发

### 11.3.1 时间序列预测模型

### 11.3.2 异常检测算法实现

### 11.3.3 多维数据关联分析

## 11.4 自然语言生成报告

### 11.4.1 数据洞察抽取技术

### 11.4.2 报告文本生成模型训练

### 11.4.3 个性化报告定制功能

## 11.5 交互式报表与数据探索

### 11.5.1 动态报表设计与实现

### 11.5.2 自助式数据分析工具开发

### 11.5.3 语音交互式数据查询系统

# 第12章: 智能质检系统开发案例

## 12.1 计算机视觉在质检中的应用

### 12.1.1 图像分类在质检中的应用

### 12.1.2 目标检测技术在缺陷识别中的运用

### 12.1.3 图像分割在精细质检中的实践

## 12.2 缺陷检测模型训练

### 12.2.1 数据采集与标注策略

### 12.2.2 模型选择与架构设计

### 12.2.3 小样本学习在缺陷检测中的应用

## 12.3 实时视频流处理技术

### 12.3.1 视频流预处理与帧提取

### 12.3.2 并行处理pipeline设计

### 12.3.3 实时分析结果可视化

## 12.4 边缘计算在质检中的应用

### 12.4.1 模型轻量化与量化技术

### 12.4.2 边缘设备选型与部署

### 12.4.3 边缘-云协同处理架构

## 12.5 质检系统与MES集成

### 12.5.1 质检结果数据格式设计

### 12.5.2 与MES系统的接口开发

### 12.5.3 质检数据分析与报告生成

# 第13章: 智能风控系统开发案例

## 13.1 金融风险评估模型设计

### 13.1.1 信用评分模型开发

### 13.1.2 欺诈检测模型实现

### 13.1.3 市场风险量化模型

## 13.2 异常交易检测算法实现

### 13.2.1 时间序列异常检测方法

### 13.2.2 图分析在关联交易检测中的应用

### 13.2.3 深度学习在复杂模式识别中的运用

## 13.3 实时风控决策引擎开发

### 13.3.1 规则引擎设计与实现

### 13.3.2 机器学习模型与规则的融合

### 13.3.3 实时计算框架选择与应用

## 13.4 模型解释性与合规性保障

### 13.4.1 模型解释技术(SHAP, LIME等)

### 13.4.2 模型公平性评估与优化

### 13.4.3 风控模型审计与报告生成

## 13.5 系统监控与应急响应机制

### 13.5.1 风控指标监控dashboard设计

### 13.5.2 风险预警机制实现

### 13.5.3 应急处理流程与自动化响应

# 第四部分: AI大模型应用优化与运维

# 第14章: 模型优化与性能调优

## 14.1 模型压缩与量化技术

### 14.1.1 模型剪枝方法

### 14.1.2 知识蒸馏技术

### 14.1.3 低精度量化实现

## 14.2 分布式训练与推理加速

### 14.2.1 数据并行与模型并行策略

### 14.2.2 梯度累积与混合精度训练

### 14.2.3 模型服务化与负载均衡

## 14.3 AutoML在模型优化中的应用

### 14.3.1 神经架构搜索(NAS)技术

### 14.3.2 超参数优化方法

### 14.3.3 AutoML平台搭建与应用

# 第15章: AI系统安全与隐私保护

## 15.1 AI模型安全威胁分析

### 15.1.1 模型逆向与知识产权保护

### 15.1.2 对抗性攻击类型与防御

### 15.1.3 数据投毒攻击检测

## 15.2 对抗样本防御策略

### 15.2.1 对抗训练技术

### 15.2.2 输入净化与特征压缩

### 15.2.3 集成防御方法

## 15.3 联邦学习在隐私保护中的应用

### 15.3.1 横向联邦学习实现

### 15.3.2 纵向联邦学习架构设计

### 15.3.3 安全多方计算在联邦学习中的应用

# 第16章: AI应用监控与运维

## 16.1 AI系统监控指标设计

### 16.1.1 模型性能指标定义

### 16.1.2 系统资源使用监控

### 16.1.3 业务KPI与AI指标的关联分析

## 16.2 模型drift检测与处理

### 16.2.1 数据drift检测方法

### 16.2.2 概念drift识别技术

### 16.2.3 自动模型更新策略

## 16.3 自动化运维与故障恢复

### 16.3.1 CI/CD pipeline for AI

### 16.3.2 模型版本控制与回滚机制

### 16.3.3 自愈系统设计与实现

# 第五部分: AI大模型应用前沿与展望

# 第17章: 新兴AI技术在企业中的应用

## 17.1 强化学习在业务优化中的应用

### 17.1.1 库存管理优化案例

### 17.1.2 动态定价策略实现

### 17.1.3 资源调度问题求解

## 17.2 图神经网络在关系分析中的应用

### 17.2.1 社交网络分析

### 17.2.2 金融风控中的图分析

### 17.2.3 推荐系统中的图表示学习

## 17.3 元学习在快速适应场景中的应用

### 17.3.1 小样本学习技术

### 17.3.2 终身学习系统设计

### 17.3.3 快速领域适应方法

# 第18章: AI伦理与未来趋势

## 18.1 AI伦理准则与实践

### 18.1.1 AI伦理框架设计

### 18.1.2 算法公平性评估方法

### 18.1.3 AI透明度与可解释性实践

## 18.2 AI监管与合规要求

### 18.2.1 国内外AI监管政策解读

### 18.2.2 AI系统合规性评估

### 18.2.3 隐私计算技术在合规中的应用

## 18.3 企业AI应用未来展望

### 18.3.1 AGI发展

### 18.3.2 AI与传统行业深度融合趋势

### 18.3.3 AI驱动的商业模式创新

# 附录 A: AI 大模型开发工具与资源

这些附录章节涵盖了广泛的补充材料，可以帮助读者更好地理解和应用书中的内容：

附录 A 提供了实用的开发工具和资源列表。

附录 B 复习了必要的数学基础。

附录 C 关注法规和标准，有助于合规开发。

附录 D 直接链接到实践案例的代码。

附录 E 为想深入研究的读者提供了进一步阅读的材料。

附录 F 提供了实用的最佳实践清单。

附录 G 则是一个全面的术语表，有助于读者理解专业词汇。

### A.1 主流深度学习框架对比

### A.1.1 TensorFlow

### A.1.2 PyTorch

### A.1.3 JAX

### A.1.4 其他框架简介

### A.2 NLP 工具包与预训练模型

### A.2.1 Hugging Face Transformers

### A.2.2 spaCy

### A.2.3 NLTK

### A.2.4 常用预训练模型列表

### A.3 数据处理与可视化工具

### A.3.1 Pandas & NumPy

### A.3.2 Matplotlib & Seaborn

### A.3.3 Plotly & Dash

### A.4 模型部署与服务工具

### A.4.1 Docker & Kubernetes

### A.4.2 TensorFlow Serving

### A.4.3 ONNX Runtime

### A.4.4 Triton Inference Server

# 附录 B: 常用数学与统计知识回顾

### B.1 线性代数基础

### B.2 概率论与统计学要点

### B.3 优化算法概览

### B.4 信息论基础

# 附录 C: AI 相关法规与标准

### C.1 国内 AI 相关法规汇总

### C.2 国际 AI 伦理与监管趋势

### C.3 行业特定 AI 应用标准

# 附录 D: 案例研究代码仓库

### D.1 代码仓库使用说明

### D.2 环境配置指南

### D.3 各章节案例代码链接

# 附录 E: AI 大模型相关论文推荐

### E.1 基础模型架构相关论文

### E.2 领域特定应用相关论文

### E.3 模型优化与部署相关论文

# 附录 F: AI 大模型应用开发最佳实践清单

### F.1 项目规划与需求分析清单

### F.2 数据准备与预处理清单

### F.3 模型开发与训练清单

### F.4 系统集成与部署清单

### F.5 监控与维护清单

# 附录 G: 术语表

### G.1 AI 与机器学习基础术语

### G.2 深度学习与神经网络术语

### G.3 自然语言处理专业术语

### G.4 企业级 AI 应用相关术语

# 附录 H：软件 2.0 参考资源

### H.1 软件 2.0 相关学术论文

### H.2 行业报告与白皮书

### H.3 软件 2.0 社区与开源项目