

蔡天放

caiyl1999@proton.me
18741825880



教育背景

西安电子科技大学，电子科学与技术，本科	2018.9 - 2022.7
• 担任职务：行政班班长、专业班班长、书院助理	
燕山大学，人工智能，硕士	2023.9 - 2026.7
• 研究方向：图像处理、PINN(物理信息神经网络)、智能体开发等	

专业技能

编程语言：熟悉 Python、了解 C++
深度学习框架：熟悉 PyTorch 框架，能够进行模型设计和分布式训练。了解 CUDA
数据库：MySQL
Agent 开发：熟悉基于 LangChain 和 LangGraph 的 Agent 开发流程，能够结合 ChromaDB 实现向量存储与检索，利用 LlamaIndex 进行数据索引和查询优化。了解 Agent 开发中的模块化设计、状态管理以及多 Agent 协同工作机制。
大模型技术
• 部署与量化：熟悉 vLLM、Ollama、LMDeploy 等推理部署工具，了解模型量化和压缩技术，能够实现大模型的高效推理和服务化部署。
• 微调框架：熟悉 Llama-Factory、XTuner 等微调工具，能够使用这些框架对大模型进行领域适配和性能优化。
• 微调方法：掌握 LoRA、QLoRA 等参数高效微调方法，能够在有限计算资源下实现大模型的高效微调。
• 了解大模型推理框架与多智能体技术 (A2A)：了解大模型推理中的关键技术，如模型上下文协议 (MCP)、自检索增强生成 (self-RAG)，能够应用这些技术提升模型推理的准确性、效率以及复杂任务的分解与协同解决能力。

其他技能：熟悉 Linux 常用命令、Docker 容器化部署

实习经历

美的-中央研究院-AI4S 实习生	2025.5 - 2025.8
• 描述：空调噪声的测量成本极其昂贵,使用 AI 算法结合大模型 Agent 来建立传感器与噪声之间的关系以减少测量成本,并搭建大模型 Agent 框架自动化解决类似工业界数据分析	
• 我的职责：1. 基于迁移学习理论，针对工业实际生产的数据 (高纬度, 数据量少, 有噪声) 的难点, 使用 di-PLS 算法 (PLS 算法结合迁移学习) 以及数据的预处理及清洗, 实现了从加速度传感器-> 噪声的 AI 回归模型, 误差在 3dB 以内的精度为 99%(行业内领先)2. 针对此类问题以及工业界普遍需求, 开发了基于 LangGraph 下的 OpenEvolve 程序, 利用大模型多智能体架构, 基于遗传算法全自动进行代码优化/数据分析, 并添加了 RAG 系统来解决工业界内部问题复杂度高, 专业度强的问题	
• 成果：1. 实现了从加速度传感器-> 噪声的 AI 回归模型, 误差在 3dB 以内的精度为 99%(行业内领先)2. 开发了 OpenEvolve 在 LangGraph 框架下的 Demo，以此次数据案例为例，在每个岛屿迭代 200 次的情况下，可以实现 3dB 以内 95% 的精度，接近人工处理水平 3. 调查了 AI4S 的行业内进展以及未来的发展趋势	

项目经历

Openevolve-LangGraph https://github.com/caiyu1999/qiuzhao_ai4s_agent	2025.5 - 2025.8
• 描述：本项目基于 Google 公司 AlphaEvolve, 通过结合大型语言模型 (LLMs) 的创造力和自动化评估机制，自主发现、优化算法并解决复杂科学问题。	
• 我的职责：主要工作内容: 使用 Langchain 以及 LangGraph 的框架重构并修改了整个 workflow, 整个算法基于遗传算法, 利用 LangGraph 的子图并行机制，在多个岛屿上进行代码的演化, 并定期进行岛屿之间代码的迁移，利用提示词模板以及结构化输出获得大模型输出后的增强 (变异) 代码作为子代，并随着不断的演化增加代码的精度。(精度可自定义, 如速度，数值精度等)，为特定的专业性问题添加了 RAG 架构，可根据需求添加专业领域的知识来辅助代码演化。	
• 成果：在实习期间利用此 demo 在真实的工业数据上进行了模拟，以解决复杂工业数据分析问题，最终精度与人工手动处理相差 5%(每个岛屿迭代 200 次)，在加入了部分人工筛选的 RAG 文档后，每个岛屿迭代 130 次即可达到相同的精度，极大的降低了成本，提高了演化的效率。	

- 描述：本项目基于 LangGraph、Streamlit、Llama-Index、Chroma 向量数据库和微调大模型，开发了”基于 Self-RAG 的智能劳动法咨询助手”系统。系统能够自动加载、解析和索引《中华人民共和国合同法》等权威法律条文，实现用户自然语言提问、法律条文检索、重排序、智能问答及依据展示等功能。
- 我的职责：主要工作内容：负责项目整体架构设计，基于 LangGraph 的 Self-RAG 架构，设计包含检索、评估、生成、重写、质量评估五个核心节点的智能 workflow，智能回溯节点，重写模糊提问，评估文档相关性，采用 Streamlit 快速搭建交互式 Web 前端，实现用户与法律知识库的智能对话。利用 Llama-Index 框架，完成法律条文的向量化存储与高效检索。使用 LLamafactory 微调 Qwen1.5B 模型，规范回答格式，提升回答准确度。实现法律问答的自动生成，并支持思维链展示，提升模型可解释性。实现法律条文的自动加载、结构化验证和节点化处理，保障数据一致性与检索准确性。引入 Rerank 重排序模型对检索结果进行重排序，显著提升答案相关性和权威性。支持依据条文的可视化展示，方便用户追溯法律出处。优化系统性能，支持知识库的持久化存储与快速初始化，提升用户体验。
- 成果：项目成果：成功上线智能劳动法咨询助手原型，支持多轮问答、依据溯源和模型思考过程展示,在召回率上达到 93% 的准确度。显著提升劳动法咨询自动化水平，为用户提供权威、可追溯的法律解答。

获奖经历

华为杯数模竞赛三等奖，队长

2024

- 选题：风电场有功功率优化分配
- 负责算法设计和论文撰写
- 创新性地设计了基于小波变换和四点法的实时监控优化方法
- 实现了电池充放电的实时调节，优化了功率分配策略

MathorCup 大数据挑战赛三等奖，队长

2025

- 选题：“平移置换”搬迁优化决策支持模型
- 负责算法设计
- 创新性地设计了基于图结构的房屋搬迁优化方法
- 实现了自动化的补偿策略、空间整合与效益分析

英语水平

六级