

刘善坤  
电话: 18326119875 | 邮箱: [873550310@qq.com](mailto:873550310@qq.com)  
求职方向: AI 大模型、数据、图形、算法



## 工作经历

上海柏楚电子科技股份有限公司-算法实习生-软件研发部

2024.10-2025.02

- 进行排样算法的研究实验和落地应用
- 独立寻找阅读英文文献资料，独立复现论文算法，完成报告并向上汇报
- 独立完成新算法的设计和测试，输出对应报告文档
- 积极组内讨论，提出意见和方法，并主动实施落地新方法

## 教育经历

桂林电子科技大学-计算机科学与技术-全日制统考本科-计算机与信息安全学院

2022.07-2026.07

- 主要课程: 程序设计与算法、机器学习、数据结构、计算机图形
- 主要荣誉/奖项:
  - 中国大学生程序设计大赛国家级铜牌
  - 蓝桥杯程序设计大赛全国总决赛三等奖

## 项目经历

**Langgrapg 架构设计-Auto-Feature Engineering 自动特征工程因子挖掘**

核心技术: **Langgraph、LLM、Agent** 系统设计、**Agent** 架构设计、**Self-Correction**、**EDA** 数据分析、**prompt Engineering**

主要内容:

- 核心架构: 基于 LangGraph 设计多节点协同工作流 (StateGraph)，定义 AgentState 状态机管理历史尝试记录、代码片段及评分指标，实现非线性研发流程的系统化管理。
- 智能决策: 采用 ChatOpenAI 作为核心控制器，设计精细化 Prompt 引导 LLM 结合农业/数据科学背景（如 NPK 元素平衡逻辑）提出创新特征想法。
- 闭环自纠错: 开发 Code Executor 模块，利用 exec() 动态执行生成的 pandas 代码；引入 Do-While 纠错循环，当代码运行报错（如拼写错误、逻辑异常）时，Agent 能结合 traceback 错误信息自动进行多轮代码修复（Retry Logic）。
- 效能评估: 集成自定义 API 验证模块，对 75 万行规模的数据集进行实时 MAP@3 分数评估，并将反馈结果 (Score) 持久化至实验日志，形成“思考-实践-反馈”的持续进化闭环。
- 复杂工作流控制: 使用了 add\_conditional\_edges (条件边) 来控制 Agent 的迭代次数（如 20 轮迭代后自动停止），可以对 LangGraph 流程控制 的深度掌握。

**Pytorch 模型设计-基于量子动力学启发的时间序列连续概率密度演化研究（实验性项目）**

核心技术: **Pytorch、KDE、Transfromer、量子物理、数值计算、模型架构设计**

核心研究内容:

- 数学框架构建: 尝试将特征空间定义为波函数的测度空间，引入含时薛定谔方程作为演化算子，利用其幺正演化 (Unitary Evolution) 特性探索时序特征在长距离传递中的信息守恒问题。
- PINNs 架构实现: 基于 Physics-Informed Neural Networks 思想，利用自动微分技术计算拉普拉斯算子和时间导数，构建了包含动力学残差、概率归一化和观测对齐的联合损失函数。
- 连续时间建模: 采用时间作为连续坐标输入，避免了传统模型对固定时间步长的依赖，为处理高频不规则采样数据提供了理论上的统一框架。
- 势能函数学习: 将哈密顿算符中的势能项设计为可学习的神经网络，尝试从原始数据中反演特征空间的隐式能量场。

**大模型开发与集成-基于 Gemma 3 与 集成学习的蘑菇毒性多模态预测系统**

核心技术: **Pandas、量化模型、多模态提示词工程、XGBoost、LightGBM、OOF Stacking、特征工程、Python**

项目描述: 开发了一套混合架构 (Hybrid Architecture) 预测系统。该系统利用大语言模型 (LLM) 强大的视觉语义理解能力提取高阶特征，并结合经典集成学习算法进行高精度决策，实现了从原始图像到毒性判定结果的端到端自动化流程。

主要职责与技术实现:

- 多模态语义特征提取 (LLM Layer):  
部署 Gemma 3 4B 本地量化版，设计并优化多轮 Prompt Engineering，引导模型结构化描述（如菌盖纹理、菌褶密度等）。
- 双层集成模型架构 (ML Layer):  
构建以 XGBoost 和 LightGBM 为基分类器的第一层预测模型，利用 OOF (Out-of-Fold) 策略生成元特征，有效防止训练过程中的过拟合。设计 OOF Meta Model (元模型) 进行二阶融合 (Stacking)，通过学习各基模型的预测偏差，进一步提升系统在边缘样本上的泛化性能。