



## 教育背景

至今	四川大学 (985) · 计算机学院
2023.09	计算机科学与技术 · 硕士
2023.09	四川大学 (985) · 计算机学院
2019.09	计算机科学与技术 · 学士

## 荣誉

- 2024 年英特尔奖学金
- 四川大学 2023-2024 学年优秀研究生干部
- 四川大学 2023-2024 学年优秀研究生
- 四川大学研究生一等奖学金
- 四川大学优秀学生

## 实习经历

2025.3	百度智能云 (成都)-AI 能力引擎研发组-后端研发工程师
2024.12	<ul style="list-style-type: none"><li>项目背景: TextMind 是基于文心一言 4.5 和 DeepSeek-R1 的一站式文档处理智能平台, 业务涵盖了文档抽取, 合同审查, 智慧招采。应用中心负责文档业务的具体实现, 数据中心负责用户标注数据及数据集管理, 模型中心负责用户个性化微调的 AI 原生服务文档平台。</li><li>应用中心模块, 参与<b>文档抽取</b>, <b>智慧招采</b>业务的 LLM 工作流具体设计和实现。<ul style="list-style-type: none"><li>基于 <b>Asynq</b> 分布式任务队列异步处理框架实现 <b>workflow</b> 多阶段业务解耦。灵活利用 asynq 消息聚合能力, 实现多个大模型任务聚合后并行请求 LLM 后端, 充分利用模型并行推理能力。</li><li>使用 OpenTelemetry 和 ThreadLocal, 收集分布式追踪数据, 定位链路性能瓶颈。</li><li>为了解决当实例扩容时, 资源利用率较低的情况, 对 Asynq 并发分配方式由一个协程串行打平为每个不同的 worker 协程<b>并发分配分离</b>, 可灵活进行并发控制, 响应时间减少 20% 以上。</li><li>为了解决系统吞吐量过小导致的任务超时问题, 微调文心 Speed 实现<b>多切片分类</b>, 分类 F1score 达到 0.75, 与单切片性能相当, 但服务整体速度提升了 3.3 倍。</li></ul></li><li>数据中心模块, 参与<b>智慧招采</b>业务数据飞轮, 用户通过标注数据维护数据集, 联动模型中心提升效果。<ul style="list-style-type: none"><li>用户复制或新增数据集版本时, 发起后台异步任务复制数据集版本-异步任务信息-用户标注记录, 实现版本间解耦。</li><li>大数据集统计标注信息时, 通过分页查询 Mysql 进行内存优化, 并使用游标法, 防止深分页问题。</li></ul></li></ul>
2024.11	曦谋决策 (杭州) 智能科技有限责任公司-研发部-算法开发工程师
2024.8	<ul style="list-style-type: none"><li><b>国家电网光明电力大模型-分布式光伏时序预测大模型</b>: 1. 基于时序大模型Time-LLM架构, 开发自然语言模型与时序模态对齐算法, 结合高精度天气预报数据, 构建以百度千帆大模型为核心的分布式光伏电力预测系统, 完成客户电站功率建模及超长期预报。</li><li>2. 基于 BentoML 搭建时序大模型在线推理服务, 针对用户数据特征进行个性化微调以提升预测精度, 设计微调参数存储策略, 并实现在华为昇腾 910A 计算卡上的推理部署。</li><li>3. 在河北地区实现户级 (5 万户) 48 小时及县区级 (109 个县区) 240 小时功率预测, 准确率超过 93%, 系统部署于中国电力科学研究院电力自动化所。</li></ul>
2024.08	清华大学启元实验室
2024.06	<ul style="list-style-type: none"><li>1. 基于 SAM 模型处理任务数据集, 融合 Refcoco 通用场景数据生成面向特定任务的空间感知数据集, 采用 DeepSpeed 加速训练 MiniCPM-V2.5 的空间视觉理解能力。</li><li>2. 基于 LangChain 实现多模态模型工具调用与任务规划, 使具身智能机器人能根据自然语言指令输出目标物体 2D 坐标, 完成物品抓取及状态驱动的动作指令生成。</li></ul>

## 🔧 技能

<b>Go</b>	Go 基础、协程、Gin、Gorm
<b>其他语言</b>	Python、Java（了解）、C++（了解）
<b>数据库</b>	MySQL、Redis、ElasticSearch
<b>消息中间件</b>	Asynq
<b>工具、部署</b>	Jaeger、k8s、Docker、Nginx
<b>AI 工具</b>	Pytorch、Ollama、LangChain
<b>人工智能</b>	多模态模型，时间序列-时空数据神经网络
<b>🇨🇳 语言</b>	英语 – CET6（516）

## 🌸 主要开源项目

- ▶ NuwaTS: 基于大语言模型的缺失时序数据补全基础模型 (55 stars)
- ▶ SUMformer: 城市时空流量预测方法 (16 stars)

## 🌸 科研成果

- ▶ Rethinking Urban Mobility Prediction: A Super-Multivariate Time Series Forecasting Approach TITS(SCI 一区, CCF-B)
- ▶ NuwaTS: a Foundation Model Mending Every Incomplete Time Series TNLS 在投