

王 劭

(+86)135-0587-3867 | shaowang@sjtu.edu.cn | github.com/Electronic-Waste

教育背景

上海交通大学 计算机科学与技术 学术型硕士	2024.09 - 2027.03
上海交通大学 软件工程 工学学士	2020.09 - 2024.06
• GPA: 3.76/4.30(22/99)	
• 获得过多次校级学业奖学金、专项奖学金、校三好学生等各类奖学金和荣誉称号	

开源贡献

Kubeflow WG AutoML/Training Maintainer	
• LLMOps: 主导 Kubeflow LLM Trainer V2 的设计和实现, 提供预配置蓝图, 支持 <i>LoRA/QLoRA/DoRA</i> 以及全量微调策略, 简化 Kubernetes 上大模型的微调流程, 优化用户体验	
• Job Orchestration: 与 Kubernetes 社区合作, 根据 Kubeflow LLM Trainer V2 的需求改进 JobSet 组件	
• Mentorship: 担任多个 GSoC'25 项目的 Primary Mentor, 指导帮助学生完成项目	
• 项目维护: 负责 <i>kubeflow/trainer</i> 和 <i>kubeflow/katib</i> 组件的日常维护, 参与 PR Review 和 Issue 解决	
(其它开源贡献包括但不限于 <i>vllm/production-stack</i> , <i>kubeedge/sedna</i> , <i>tensorchord/envd</i>)	

实习经历

Databend Labs 云平台实习生	2024.10 - 2025.04
• Operator: 从 0 到 1 搭建 Databend 的开源 Operator, 支持磁盘缓存、滚动升级、日志采集、Service/Ingress 等功能, 为开源用户提供灵活高效的 Databend 集群编排能力	
• 可观测性: 在 Databend 私有云平台中引入了 Prometheus 监控模块, 提供指标查询接口, 可视化系统运行状态	

谷歌开源之夏 学生贡献者, CNCF - Kubeflow	2024.05 - 2024.09
• Push-based 指标采集: 实现了一套基于主动推送模式的机器学习训练指标采集方案, 取代 Katib 原有的基于 Sidecar 的训练数据采集模式, 降低了性能/资源开销和指标采集的时延	
• Python SDK: 为数据科学家实现了一个简洁的训练数据推送接口, 取代原有基于标准输出/日志文件的指标输出方案, 降低指标采集过程中输出格式出错的可能性, 增强系统的鲁棒性	
• e2e 测试: 在 Github Actions 框架下为超参数调优函数编写 e2e 测试集, 以提升 CI 的完备性	

上海邦戴数字科技有限公司 基础架构实习生	2024.01 - 2024.04
• 性能测试: 针对期货期权和股票行情数据分析场景 (OLAP) 调研主流时序数据库, 给出性能测试报告	
• 行情数据存储系统: 从 0 到 1 在 Kubernetes 集群上设计搭建新的行情存储系统, 取代现有基于分布式文件系统和 csv 的行情数据存储系统, 将数据分析业务的性能提升 10x 倍以上	

技能

- 编程语言: Golang, Python, C/C++, Java
- 英语: CET6 Score: 605/710