

# 李逸凡

18811019056 | liyifan.steven616@gmail.com

## 教育经历

### 宾夕法尼亚大学

2021 年 9 月 - 2023 年 5 月

工学硕士：数据科学，工程与应用科学学院

美国 - 费城

- A+ 课程：Algorithms & Computation, Machine Learning, Deep Learning, Artificial Intelligence, Computational Linguistics

### 清华大学

2017 年 9 月 - 2021 年 6 月

管理学学士：信息管理与信息系统，经济与管理学院

中国 - 北京

辅修学位：数据科学与技术，软件学院

- 平均分：3.84/4.0
- 获 2019 年国家奖学金、2018 & 2020 年学业优秀奖学金等多项荣誉

## 工作经历

### 亚马逊

2024 年 3 月 - 至今

软件开发工程师

美国 - 西雅图

- 从零设计并搭建团队端到端数据流水线与机器学习平台，实现每天约 70GB 数据处理能力，用于扫描亚马逊内部数据库与 API 并进行基于元数据的分类；支撑下游针对高数据安全严重性与个人数据保护策略的执行
- 架构并实现一套 AWS 原生流水线：使用 Step Functions 进行调度, AWS Glue (Scala) 做 ETL, SageMaker 覆盖模型全生命周期 (训练、评估、部署、推理), Amazon Bedrock 用于大模型推理；通过 SQS 与 Lambda 实现可扩展的 Bedrock 作业调度，管理并发、重试与运行可靠性

### 亚马逊

2023 年 7 月 - 2024 年 3 月

软件开发工程师

美国 - 西雅图

- 使用 Scala-Spark 维护团队的 AWS Glue 数据流水线，保障系统平稳运行与数据一致性；设计并实现流水线中的新“员工消歧 (Employee Disambiguation)”组件，使转化率指标准确度提升 7%
- 开发并维护网站门户：后端采用 AWS Lambda 与 DynamoDB, 前端使用 React; 设计并实现一个“边界框 (bounding box) 绘制工具”的后端与前端组件，使用户能够定义关注区域并获取定制化指标

### 微软亚洲研究院 (MSRA)

2020 年 8 月 - 2021 年 3 月

软件分析组，实习研究员

中国 - 北京

- 以第三作者身份在 SIGKDD 2021 发表论文 *HALO: Hierarchy-aware Fault Localization for Cloud Systems*
- 与团队成员共同设计并实现“HALO”——用于云服务的大规模故障诊断算法；复现两项相关工作，并使用 Python 对准确率与运行时效率进行对比实验
- 开发 Flask API 服务与 Kusto 插件，便于其他内部团队使用我们的新算法；根据 Office 365 与 Azure 团队成员反馈，该算法显著缩短了工程师排查异常的时间
- 实习结束时获得明日之星卓越奖

### 字节跳动

2020 年 1 月 - 2020 年 8 月

大数据研发 / 数据科学实习生

中国 - 北京

- 在公司广告数据仓库的建设与治理中应用 Hive SQL 与 Spark
- 维护两张与 TikTok 广告产品“Dou+”相关的数据集市 (data mart) 级别表的表结构与数据流水线；实现新增需求，例如增加与优惠券、购物车、直播推广相关的新指标
- 优化公司线上实验平台的 A/B 测试异常检测机制；调研并落地一种基于时间序列数据挖掘的新算法，使告警准确率提升 30%

## 研究经历

### 研究助理

2021 年 9 月 - 2023 年 5 月

导师：宾夕法尼亚大学沃顿商学院 Prof. Ryan Dew

美国 - 宾夕法尼亚

- 为数据驱动的营销研究构建偏好测量 (preference measurement) Web 应用；对遗留的 Flask 与 React 代码进行大量调试、重构与优化，显著提升代码健壮性与可读性
- 协助维护多个项目的数据 ETL 流水线，使用 Python、MongoDB 与 AWS S3

# 其他项目经历

<b>CNKI-2-BibTeX</b>	2019 年 2 月 - 至今
开源项目作者	
<ul style="list-style-type: none"><li>构建开源库 CNKI-2-BibTeX, 用于将中文学术数据库 CNKI 的文献信息转换为可用于 BibTeX 的格式</li><li>该项目帮助了许多中国学者开展科研工作，并获得 150+ GitHub Stars 与 10.5 万次下载</li></ul>	

# 其他信息

- 技能：Python, Scala, Kotlin, Spark, SQL, Hive, L<sup>A</sup>T<sub>E</sub>X, C, Java, TypeScript, Node.js, AWS, 数据科学，机器学习