



- 美国计算机专业 Top5 学校本科、加拿大 Top2 学校硕士
- 3 年 Java 后端开发经验，熟练掌握基于 Spring Boot 和 Spring Cloud 的微服务开发
- 长期活跃于 Java 开发者论坛 Together Java, 为初学者解答大量问题
- 具备基本的机器学习和数据分析技能，会使用 Python, R 和 AWS SageMaker

工作

2022 - 2024 加拿大	专业进修 英属哥伦比亚大学 UBC <ul style="list-style-type: none">数据科学硕士，学习统计和机器学习原理毕业实习：使用 CNN 图像模型检测 TikTok 视频是否被滤镜处理过计算机课程助教，软件开发培训 机器学习 深度学习 卷积神经网络 大语言模型 Python pandas scikit-learn PyTorch AWS SageMaker
2021 - 2021 北京	Java 软件工程师 环信/声网 Agora Inc <ul style="list-style-type: none">开发基于微服务架构和 Spring Cloud 的 SaaS 即时消息系统使用字典树和 Thrift RPC 实现了一个高性能关键词检测程序，用来过滤违规消息利用 Reactor Netty 客户端开发了一套响应式的 Java SDK，方便用户对消息平台进行管理操作为满足 GDPR 合规要求，开发了一个信封加密库对用户数据加密处理，并使用 AWS KMS 管理密钥 Spring Boot Spring Cloud MySQL Redis Tomcat Netty Reactor GitHub Actions
2020 - 2021 北京	Java 软件工程师 数语科技 <ul style="list-style-type: none">开发类似 ERwin Data Modeler 的数据治理工具，用于对大规模数据库的元数据管理完善了数据质量检测功能——系统可扫描数据库，并使用可定制的规则验证数据的合理性为实现与第三方系统的数据对接，使用 Spring MVC 开发了一套 REST API使用 JUnit 和 Mockito 为现有功能完善单元测试，使覆盖率达到 80% 以上 Spring MVC JDBC PostgreSQL Oracle MySQL Hibernate JUnit Mockito Maven
2019 - 2020 北京	全栈程序员 ShopHitly <ul style="list-style-type: none">参与早期的电商平台创业项目，向美国伊利诺伊州用户销售烟具使用 Node.js, Express.js 和 MongoDB 搭建 REST web 服务，用来管理商品库存信息用 Docker, AWS ECS 和 CodePipeline 搭建 CI/CD 管道并部署应用基于 TypeScript 和 React 搭建简单的 web UI Node.js Express.js MongoDB TypeScript React HTML CSS Docker AWS ECS REST
2018 - 2019 美国	Hadoop 数据工程师 雅虎/Verizon (全美最大的电信公司) <ul style="list-style-type: none">负责维护一个低延迟 OLAP 数据仓库系统，用于网站的用户参与分析 (user engagement)使用 Hadoop 大数据技术栈搭建 ETL 数据管道将数据导入高性能的 Apache Druid, 使用户可在几秒内对 PB 量级的数据进行聚合查询开发 Java web 服务，为查询 Druid 数据提供 REST API Hadoop Java Apache Druid Apache Pig Oozie Hive SLF4J H2 Gradle Linux

学历

2023	数据科学，硕士 英属哥伦比亚大学 UBC (世界排名第 38)
2017	计算机科学，本科 伊利诺伊大学香槟分校 UIUC (美国专业排名第 5)

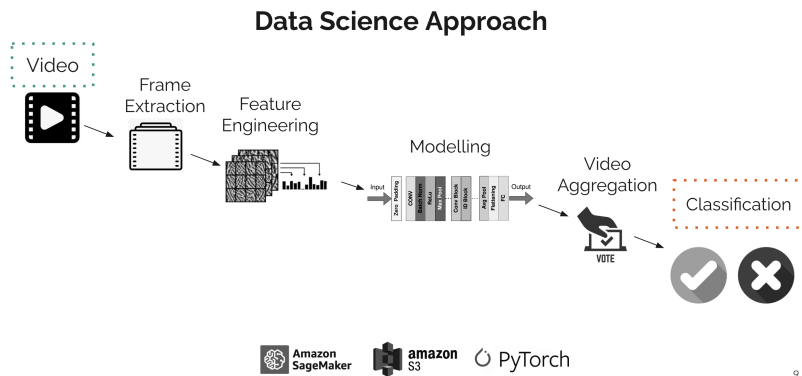
实习

2017	Java 程序员 谷歌/CDAP 数据平台
2016	Java 程序员 芝商所/CME Group (全球最大期货交易所)

短视频滤镜检测器

使用 CNN 模型训练短视频分类器，检测视频是否加过美颜滤镜

- 用 **PyTorch** 和 **scikit-learn** 搭建模型，用 **AWS SageMaker** PaaS 服务部署训练任务
- 基于预训练的 **ResNet101** 模型添加额外卷积层，再用输入图像数据做训练微调
- 对普通三通道 RGB 图像增加 **PRNU** 噪点信息通道，使模型能检测出滤镜对噪点层的影响
- 尝试多种 **CNN** 模型结构和不同的图像预处理组合，并做大量超参数调优



即时消息系统升级改造

为 IM 即时消息 SaaS 服务商公司实现新功能

- 用字典树实现消息的关键词过滤服务，与消息模块通过 **Thrift RPC** 通信
- 用 **Reactor Netty** HTTP 客户端实现了一套响应式 **Java SDK**，用来管理消息元数据
- 为满足 **GDPR** 要求，开发了一个信封加密库对用户数据加密，并使用 **AWS KMS** 管理密钥
- 改造原有的 **JWT token** 结构，以对新收购公司的用户做统一鉴权处理
- 完善现有代码中千疮百孔的单元测试，并使用 **wrt Lua** 脚本做压力测试

数据质量扫描器

在原有数据治理软件的基础上，增添自动数据质量扫描功能

- 支持多种关系型数据库，如 **Oracle**, **SQL Server**, **OceanBase** 和 **PolarDB**
- 使用 **ANTLR** 做自定义命令语法解析，使用户可用简洁的命令指定数据质量扫描规则
- 用 **Spring Schedule** 实现定时任务，用户可在网页端界面调度多任务运行
- 使用 **Spring JavaMailSender** 实现在任务状态更新时，邮件通知用户

电商平台创业

目标是向伊利诺伊州消费者销售烟具，我是3个创始程序员之一

- 从零到一搭建 **CI/CD pipeline**，实现服务端在 **AWS ECS** 上的容器化和弹性化部署
- 使用 **Node.js** 和 **Express.js** 实现用户、存货和订单管理 API
- 数据库使用 **MongoDB**，ORM 框架使用 **mongoose**
- 用 **Socket.io** 库实现网页端基于 **WebSocket** 的用户间消息通讯功能
- 集成第三方 OCR 和身份验证服务，基于驾照对用户做实名认证
- 同时负责对初级程序员的招聘，以及与 freelancer 前端开发者的对接工作