

# YIFAN LI

267-781-5871 | liyifan.steven616@gmail.com

## EDUCATION

### University of Pennsylvania

Master of Science in Engineering (MSE) in Data Science

Sep 2021 - May 2023

Philadelphia, PA

- GPA: 4.0/4.0

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

### Tsinghua University

Bachelor of Management in Information Management and Information Systems, Major

Sep 2017 - Jun 2021

Beijing, China

Bachelor of Science in Data Science and Technology, Minor

- GPA: 3.84/4.0

- Awards: 2019 National Scholarship (Top 0.01%), 2018 & 2020 Academic Excellence Scholarship

## PROFESSIONAL EXPERIENCE

### Amazon

Mar 2024 - Present

Software Development Engineer, Data Intelligence Governance Team

Seattle, WA

- Designed and built the team's end-to-end data pipeline and ML platform from the ground up, enabling daily processing of about 70GB data, to scan internal Amazon databases and APIs and perform metadata-based classification; unlocked downstream enforcement campaigns for high-severity and personal data protection policies
- Architected an AWS-native pipeline using Step Functions for orchestration, AWS Glue (Scala) for ETL, SageMaker for the full model lifecycle (training, evaluation, deployment, inference), and Amazon Bedrock for LLM inference; implemented scalable Bedrock job orchestration with SQS and Lambda to manage concurrency, retries, and operational reliability

### Amazon

Jul 2023 - Mar 2024

Software Development Engineer, Physical Store Analytics Team

Seattle, WA

- Maintained team's AWS Glue data pipeline using Scala-Spark, ensuring smooth operation and data integrity. Designed and implemented a new Employee Disambiguation job the pipeline, improving conversion rate metric accuracy by 7%
- Developed and maintained the website portal with an AWS Lambda backend and DynamoDB database, along with a React frontend; designed a bounding box drawing tool and implemented both backend and frontend component of it, enabling users to define areas of interest and obtain tailored metrics

### Microsoft

Aug 2020 - Mar 2021

Research Intern, Microsoft Research

Beijing, China

- Published research paper, *HALO: Hierarchy-aware Fault Localization for Cloud Systems*, on SIGKDD 2021 as the third author
- Designed and implemented "HALO", a large-scale failure diagnosis algorithm for cloud services, with other team members. Reproduced two related works and conducted comparative experiments of accuracy and runtime efficiency with Python
- Developed Flask API services and Kusto plugins that allow other internal teams to use our novel algorithm. According to feedback from members of Office 365 and Azure, our algorithm significantly reduced engineers' anomaly mitigation time
- Awarded the "Stars of Tomorrow" award of excellence at the end of the internship

### ByteDance

Jan 2020 - Aug 2020

Big Data R&D Intern / Data Science Intern, Ads Data Platform

Beijing, China

- Applied Hive SQL and Spark in corporate advertisement data warehouse construction and governance
- Maintained the schema and data pipeline of two data-mart-level tables related to "Dou+", the advertising product of TikTok. Implemented feature requests such as adding new metrics related to coupons, shopping carts, and live stream promotions
- Optimized the A/B tests anomaly detection mechanism on the corporate online experiments platform. Conducted research and implemented a new algorithm with time series data mining techniques that increased the alerting accuracy by 30%

## RESEARCH EXPERIENCE

### Quantitative Marketing Research Assistant

Sep 2021 - May 2023

The Wharton School, University of Pennsylvania

Philadelphia, PA

- Built a preference measurement web application for data-driven marketing research. Debugged, refactored, and optimized the legacy Flask and React code extensively, leading to significant improvement of code robustness and readability
- Assisted the maintenance of data ETL pipeline of several other projects using Python, MongoDB, and AWS S3

## OTHER PROJECTS AND ACTIVITIES

### CNKI-2-BibTeX

Feb 2019 - Present

Open Source Developer and Maintainer

- Built an open-source library, CNKI-2-BibTeX, for using BibTeX with a Chinese academic database CNKI
- The project helped many other Chinese scholars conduct researches, and got more than 150 stars and 105k downloads

## ADDITIONAL INFORMATION

- Skills: Python, SQL, Hive, Scala, Spark, TypeScript, Node.js, Data Mining, Machine Learning, AWS