

# Zhihao Lyu

6694997073

zhly4336@colorado.edu

[Personal Website](#)

## EDUCATION

### University of Colorado Boulder

*Doctor of Philosophy in Computer Science, GPA: 3.7/4.0*

Sep. 2024 ~ Jun. 2028

Boulder, CO

### Northeastern University

*Master of Science in Computer Science, GPA: 3.9/4.0*

Sep. 2020 ~ Dec. 2023

Boston, MA

### Beijing University of Civil Engineering & Architecture

*Bachelor of Engineering in Urban Planning, GPA: 3.5/4.0*

Sep. 2015 ~ Jun. 2020

Beijing, China

## EXPERIENCE

### Tripo AI

*Software Engineer*

Feb. 2024 ~ Aug. 2024

Beijing, China

- Developed public APIs and web backend, supported 2 major version launches. Led the migration from Google Analytics to **Posthog** to enhance user tracking for the operations team, resulting in an increase in DAU/WAU from **2k/20k to 4k/30k**.
- Set up a delay queue using **Redis** to manage low-priority service calls, effectively reducing server load. Implemented distributed locks to prevent duplicate consumption issues across multiple workers.
- Implemented a **Kafka** client in **Go Fiber** with goroutines and semaphores for handling callbacks from the training cluster. Designed an idempotent key to prevent duplicate consumption within a same consumer group.
- Refactored asynchronous Redis, PostgreSQL, and Kafka clients using **Python FastAPI**, and utilized Lua scripts and pipelines in Redis to estimate queueing time and task completion time, significantly improving user experience.
- Used topological sort to convert inference pipelines into **DAGs** to enable concurrent task execution, **reducing runtime from 50s to 20s**. Introduced Temporal as the task scheduler for retrying failed tasks and periodic backups from Redis to databases.

### MIT Transit Lab

*Research Assistant*

Aug. 2023 ~ Dec. 2023

Boston, MA

- Collaborated with Chicago Transit Authority, funded by the Department of Energy, to address public transportation service reliability issues, employing machine learning algorithms for bus scheduling.
- Established a cloud computing platform using Flask, modified distributed transaction for database, and integrated Google Cloud Platform to store and continuously update bus location, speed, and arrival time.
- Utilized Data Frame Algebra in Pandas to merge these real-time data and calculate service reliability metrics such as number of passengers, waiting time, and cycle time. Updated bus scheduling strategy every minute, and offered a web application for experts to evaluate scheduling strategy using React.js.
- Successfully operated the system for three researchers and dozens of dispatchers for months, collecting over 10,000 high-quality data points to improve algorithm performance.

### Sleep Number Lab

*Cloud Engineer Intern*

May 2022 ~ Aug. 2022

San Jose, CA

- Assisted the Cloud team (AWS-based) in building a distributed data platform to capture 1 billion daily bio-data using Spring and Kafka. Implemented a real-time data pipeline to transform data from CSV to Parquet for the machine learning team.
- Optimized a Lambda function with multiprocessing techniques, resulting in a 30% cost reduction and reducing the average processing time within the pipeline by 46%.
- Utilized SQL-like queries to extract and store data into DynamoDB and S3, triggered through EventBridge.
- Migrated MQTT brokers from self-managed servers to SaaS and developed a script for over 1 million IoT devices using Python for seamless provision in IoT Core through the usage of Cognito and IAM.

## SKILLS

**Languages:** Java, Python, Go, SQL, C, Typescript, HTML/CSS

**Frames & Tools:** Spring, Flask, FastAPI, gRPC, Fiber, React, Node, Express, Kafka, MyBatis, Docker, AWS, GCP