

# Chenggang Li

(+86)18640056117

[me@chenggang.li](mailto:me@chenggang.li)

## EXPERIENCE

---

### TikTok

Beijing, China

Software Engineer 2

Jan. 2024 – Present

Worked on Data Synthesis Pipelines and backend service development at AIDP(AI Data Platform) Team

- Worked on multi-round LLM Unit Test Data Synthesis Pipelines for SFT/RL finetuning Doubao code models.
- Designed and implemented multiple backend microservices using Python/Golang/Kafka/Redis/MySQL stack.
- Implemented and Carried out intelligent distribution experiments on different task assignment strategies such as similar tasks clustering, MILP based global optimum assignment, LLM-assisted pre-labeling, etc.
- Integrated Vector Database into existing architecture to push semantic-related task based on vector similarity to reduce labeler context switching, reduced average handling time by 12%. (Patent Pending).
- Increased labeler efficiency by 10% after implementing novel task-to-labeler distribution strategy based on labeler's profile and historical data. Evaluated gains by analyzing labeler's performance data with Hive SQL.

### Microsoft

Beijing, China

Software Engineer

Sep. 2021 – Dec. 2023

Designed and implemented video index generation services and ETLs at Bing Multimedia Video Index Generation Team.

- Optimized Bing Video Search Index Freshness to match and beat Google's and Baidu in side-by-side metrics.
- Built and maintained real-time and batch process ETL pipelines for video discovery, ingestion and transformation.
- Improved index size from 120M to 300M by code profiling and optimization without additional capacity ask. Increased index generation service QPS by 3x which mitigated our cluster's Kafka consumer lag issue.
- Developed monitoring solutions to ensure performance health and reliability for ETL and services.
- Discovered, identified and fixed a high impact memory leak issue (~10GB/hour) in core service.

### eBay

San Jose, USA

SDE Intern

Jun. 2021 – Aug. 2021

- Explored query optimization methods on Nugraph, an inhouse JanusGraph based open-source graph database.

### Microsoft Research

Beijing, China

Research Intern

Feb 2017 – Feb 2018

- Proposed a Machine Learning (Random Forest) failure prediction method, reduced Azure VMs' failure rate by 43%.

## EDUCATION

---

### University of Southern California

Los Angeles, USA

Master of Science in Computer Science

Jul. 2019 – Aug 2021

- Annenberg Fellowship

### Northeastern University

Shenyang, China

Bachelor of Engineering in Computer Science

Oct. 2014 – Jun. 2019

- Dean's List for Innovation and Entrepreneurship

## AWARDS

---

- Winner of Ability Award (1/150 countries) in Imagine Cup World Final.
- Winner of Microsoft Imagine Cup China 2016.
- SIGSOFT CAPS Award, ACM SIGSOFT, Sep 2018

## PUBLICATIONS

---

Predicting Node Failure in Cloud Service Systems, Qingwei Lin et al.

- Co-author. Appeared on ESEC/FSE 2018. DOI: [10.1145/3236024.3236060](https://doi.org/10.1145/3236024.3236060)

UI Test Migration Across Mobile Platforms, S Talebipour et al.

- Co-author. Appeared on ASE 2021. DOI: [10.1109/ASE51524.2021.9678643](https://doi.org/10.1109/ASE51524.2021.9678643)