

Tony Xiao

Education

Brown University

Master of Science in Computer Sciences

Aug 2024 - May 2026

University of Illinois at Urbana-Champaign

Bachelor of Science in Computer and Information Sciences

Aug 2021 - May 2023

Coursework: Data Structure and Algorithms, Database System and Design, Computer System, Distributed System

Professional Experience

Huawei | Software Engineer

Aug 2023 - Aug 2024

- Implemented a multi-layered caching system using **Caffeine and Redis** with LRU and TTL eviction strategies, improving response time from **3s to 240ms** and enhancing system performance during high-concurrency scenarios
- Architected a real-time data processing pipeline using **Kafka** to ensure reliable transmission of large volumes of data, optimizing throughput to handle over **10,000 transactions per minute** during traffic spikes
- Constructed a scalable batch image retrieval system using **AWS SDK for S3** and **Spring Cloud's** asynchronous task execution, enabling efficient handling of up to **1000 images per batch**
- Deployed microservices on **EKS (Elastic Kubernetes Service)** using **Docker** and established CI/CD pipeline with **GitHub and Jenkins**, expediting deployment by **80%**
- Implemented comprehensive monitoring and logging solutions using **Apache SkyWalking** and **ELK stack** (Elasticsearch, Logstash, Kibana) to resolve production issues, achieving **95% reduction** in system downtime

Bashpole Software | Software Engineering Intern

Feb 2023 - May 2023

- Developed a conversion tracking system using **Google Analytics** and **Tag Manager** to monitor user engagement metrics and ROI for **30 nonprofit organizations**
- Built data-driven advertisement integrity systems that improved campaign performance by **50%** through enhanced tracking of key metrics
- Automated advertising email workflows with **Google Apps Script**, reducing processing time by **66%** and ensuring consistent data handling

UIUC | Undergraduate Research Assistant

Sep 2022 - Dec 2022

- Led development of a large-scale data processing system that handled over **2.4 million** social media posts, implementing custom machine learning models for text classification and sentiment analysis
- Built a schema-less architecture using **Python** and **AWS Lambda** for automated data scraping and processing, significantly expediting the model training procedure
- Engineered a real-time data pipeline integrating **BERTopic** and **RoBERTa** models for large-scale data classification, reaching **10,000 monthly users**

Project Experience

Vector Database Search System

April 2024 - June 2024

- Processed **110,000 products** from Amazon dataset, implementing vector embeddings for efficient semantic search
- Enhanced vector search efficiency using **FAISS**, achieving **70% improvement** in query performance through optimization of indexing strategies
- Implemented **LLM-based** response generation system with context-aware capabilities for handling user queries

Communication Networks

Feb 2022 - May 2023

- Implemented TCP protocols on UDP using **C++** and socket programming, including flow controls, congestion avoidance, and packet loss handling
- Built routing algorithms with dynamic topology adjustment capabilities, simulating real-world network behaviors

Skills

- Languages:** Java, Python, Go (Golang), JavaScript, C, C++, SQL
- Libraries & Frameworks:** React, Node.js, Spring Boot, Spring Cloud, BERTopic, RoBERTa
- Systems & Tools:** Redis, Kafka, PostgreSQL, MySQL, Firebase, Docker, Kubernetes
- Cloud Services:** AWS (S3, EC2, Lambda), Google Analytics
- Specialties:** Large-scale data processing, Caching strategies, System optimization

Availability

Available for 12-week internship starting May 27, 2025 - August 15, 2025