

# Yi (Ethan) Zeng-Tsang

📍 Boston, MA   ✉ xethts@gmail.com   in yi-zeng-927638297   🌐 eth-ts-x   👤 ethantsang.me

## Work Experience

### Research Assistant

*Boston University*

*Boston, MA*

*Mar 2024 – May 2025*

- Architected a full-stack ML prediction platform (**Next.js**, **FastAPI**) featured in a peer-reviewed [IEEE publication](#) 📄.
- Engineered an async inference pipeline utilizing **Celery** for task queuing and **Triton gRPC** for batched ONNX execution; slashed latency by over 99% against a synchronous TensorFlow baseline, driving up to a 1,100x throughput gain.
- Designed and deployed cloud infrastructure on **Azure** using **Terraform**, **AKS**, **Helm**, and **KEDA**, supporting containerized services, autoscaling workers, managed **PostgreSQL/Redis**, and secure ingress with TLS.
- Established production observability with **Prometheus**, **Grafana**, **Loki**, and **OpenTelemetry** to monitor API performance, worker health, logs, and distributed traces across the ML platform.
- Architected robust **CI/CD** pipelines via **GitHub Actions** integrating automated unit and system integration testing (**PyTest**); orchestrated zero-downtime **K8s** deployments to AKS, slashing release cycles from hours to under 7 minutes.

### Software Engineer Intern

*CLR Analytics Inc.*

*Irvine, CA*

*Jun 2024 – Oct 2024*

- Orchestrated a serverless **ETL** pipeline using **AWS Step Functions** and **Lambda** to concurrently ingest multi-agency GTFS feeds, parsing and upserting millions of transit nodes and shape geometries into PostGIS.
- Engineered low-latency geospatial workflows using **PostGIS/GeoAlchemy**; optimized complex spatial joins (**ST\_Intersects**, **ST\_DistanceSphere**) to cross-reference transit shapes with dynamic roadway detours within a 20-meter radius.
- Delivered RESTful APIs via **AWS API Gateway** to serve real-time transit disruptions, integrating both dynamic spatial queries from PostgreSQL and legacy static JSON snapshots from **S3** to ensure uninterrupted downstream consumption.

### Software Engineer

*Shenzhen KORON Soft Co., Ltd.*

*Shenzhen, Guangdong*

*Aug 2021 – Jul 2023*

- Architected a multi-tenant **SaaS** backend using **Spring Boot** and MyBatis with a **PostgreSQL database-per-tenant** model, ensuring strict data isolation and enabling independent scaling and backup strategies for high-value tenants.
- Designed a distributed event-streaming architecture utilizing **Kafka** to ingest and process 50k+ **IoT** events/sec; optimized batch consumers and **MongoDB** bulk-write operations to absorb massive traffic spikes with zero data loss.
- Resolved distributed transaction failures across 10+ microservices by implementing a **Saga** Orchestration pattern with automated compensating transactions, guaranteeing eventual consistency and reducing data anomalies by 99.9%.
- Leveraged **Grafana LGTM** Stack to monitor K8s pods and trace distributed transactions across 10+ microservices to identify bottlenecks, optimizing **PostGIS/Recursive** CTE queries to slash P99 latency by 50%.
- Resolved OOM incidents during large-scale data exports by implementing **MongoDB Cursors** and Java Stream API for flow-control processing, enabling stable export of millions of records with minimal memory footprint.

## Projects Experience

### Terrier Connect – Social Media Platform 📄

*2026*

- Architected a high-concurrency, event-driven platform utilizing **React**, a **GraphQL** gateway, and **Django** APIs, integrating **Kafka**, **Cassandra**, Postgres, Redis, and **Elasticsearch** to power scalable feeds, notifications, and search.
- Engineered a resilient, cloud-native infrastructure on **GKE** using Terraform and Helm with automated scaling, and established an end-to-end observability stack via **OTel** and **Prometheus** to monitor microservices and async pipelines.
- Architected a highly available event-driven **CDC** pipeline using **Debezium** and **Kafka** to synchronize **Cassandra** data into **ES**; eliminated distributed dual-write inconsistencies and achieved 99.9% data parity for high-scale search workflows.
- Built secure authentication flows with **JWT** in **HttpOnly** cookies and Google **OAuth**, and implemented a **Redis**-based token blacklist to enable immediate session revocation and reduce token exposure risk.

## Education

### Boston University

*Master of Science in Computer Science (GPA: 3.81)*

*Sep 2023 – May 2025*

### Nanfeng College Guangzhou

*Bachelor of Engineering in Computer Science and Technology*

*Sep 2016 – Jun 2020*

## Technologies & Skills

**Programming Languages:** Java (Proficient), Python (Proficient), SQL, Unix Bash/Shell, JavaScript, C, Go, R

**Frameworks & Tools:** Spring Boot, FastAPI, Django, React, Celery, Jenkins, Git, GraphQL, JUnit, Maven, Jira

**Cloud & Infrastructure:** AWS, Azure, GCP, Terraform, Docker, Kubernetes, Kafka, Nginx, DevOps, CI/CD

**DB & Others:** MySQL, PostgreSQL, MongoDB, Cassandra, Redis, Prometheus, Grafana, OpenTelemetry, GitHub Copilot