Jimin Yang

San Francisco Bay Area, CA | +1 (380) 257-4235 | jimin.yang.0607@gmail.com | jiminyang.com | LinkedIn | GitHub

Education

University of Illinois Urbana-Champaign **Expected Dec 2026**
Master of Computer Science **GPA:** 3.9/4.0

The Ohio State University **Aug 2024**
Bachelor of Science in Computer Information Science **GPA:** 3.85/4.0

Experience

Amazon **Jul 2026 – Sep 2026**
Software Development Engineer Intern

- Built an **AI audit workflow** on **Amazon Bedrock**: parse multimodal notices → match published rates via tool calling → verify via guardrails and human review, cutting audits from hours to **under 3 minutes** and surfacing **$20K** in dispute risk.
- Generalized the audit **AI agent** into a reusable serverless runtime on **AWS Lambda** that powers both batch audits via SQS and a live chatbot via streaming, capping turns, tokens, and time so no run can loop or overspend.
- Delivered a full-stack ocean-freight **pricing platform** used daily by **50+ sales and ops staff**, built with React/TypeScript and AWS CDK, automating ingestion, change detection, and publishing to cut price-update turnaround by **80%**.
- Shipped personalized price alerts that reach each salesperson within **1 minute** of a price release, built on **DynamoDB Streams** with a per-user outbox and **dead-letter queue (DLQ)** so failed alerts are retried, not lost.

Synopsys **Jun 2026 – Jul 2026**
Software Engineer Intern

- Developed a **regression-testing framework** for a chip-design AI agent that replays Langfuse datasets, asserts every intermediate step with pytest, and scores outputs with **LLM-as-judge**, cutting time to pinpoint failures by **70%**.
- Designed a **self-improving knowledge loop**: mine 20K+ agent chats via embeddings → LLM drafts fixes → human-in-the-loop review → **RAG** knowledge base, distilling **500+** fixes so the agent avoids repeat errors.

Blinkle AI **Dec 2025 – Mar 2026**
Full-Stack Software Engineer Intern

- Architected an **AI resume analysis pipeline** on Amazon Bedrock and SageMaker, cutting redundant inference costs by **60%** through version-aware, content-based caching and in-flight request dedup backed by Prisma transactions.
- Built **Stripe** subscription and credit billing on **NestJS** and **PostgreSQL**, processing 15+ webhook event types idempotently, with row-level locking to prevent duplicate credit grants and overdrafts.

DBS Bank **Jul 2025 – Aug 2025**
Backend Software Engineer Intern

- Added **MySQL-backed locking** to reduce duplicate processing of concurrent trade submissions in a **Java Spring Boot** microservice handling 50K+ daily trades, built with Maven and shipped via Jenkins CI/CD to Docker/Kubernetes.
- Merged PDF reporting into one **RESTful API** on MyBatis, cutting **2,000+ lines** of duplicate code and download time **60%**.

Projects

MiniBot Agent Runtime GitHub

- Implemented a local **Python agent runtime** with a single-owner turn loop, token-budgeted context compaction, and append-only session logs, preserving full interaction history while limiting model context size.
- Unified local tools and **MCP** servers behind a typed interface with approval checks and streamed run events.

CityScout GitHub

- Engineered a Java Spring Boot flash-sale ordering backend sustaining **5,000 TPS** on a single 8-vCPU VM, using **Redis Lua** scripts for atomic stock reservation and per-user purchase limits under peak contention.
- Offloaded order writes to **Apache Kafka** with idempotent consumers and DLQ recovery, cutting latency **40%**.

Skills

Languages: Java, Python, TypeScript, JavaScript, C/C++, SQL, Bash, HTML/CSS
AI & Agents: RAG, tool calling, embeddings, LLM-as-judge, MCP, Strands Agents, LangGraph, Langfuse, Bedrock, SageMaker
Backend & Web: Spring Boot, MyBatis, NestJS, Node.js, FastAPI, REST APIs, Prisma, Stripe, Redis, Kafka, React, Next.js
Data & AWS: MySQL, PostgreSQL, DynamoDB, Lambda, SQS, S3, API Gateway, EventBridge, CloudWatch, Athena, Glue, CDK
DevOps & Testing: Git, Maven, Docker, Kubernetes, Jenkins, CI/CD, CloudFormation, Linux, Nginx, JUnit, Mockito, Jest, pytest