# Mingyang Li

Fremont, CA | +1 951-966-1919 | iam.alexli371@gmail.com | github.com/mli371

**Software Engineer – Backend & AI Platform / Enterprise RAG**

## Education

**University of California, Riverside** — Riverside, CA
*M.S. in Computer Science* — *Sep. 2024 – Mar. 2026*
*Coursework: DBMS (A+), Machine Learning, Artificial Intelligence, High Performance Computing, GPU Architecture, Advanced Networks*

**Shanxi University** — Taiyuan, China
*B.E. in Software Engineering* — *Sep. 2020 – Jun. 2024*

## Technical Skills

| | |
|---|---|
| **Languages:** | Java 17, Python, SQL, C/C++, JavaScript, Go |
| **Backend:** | Spring Boot 3.x, WebFlux, Project Reactor, REST/SSE APIs, Redis, PostgreSQL / PgVector, MySQL, Kafka, Docker, MinIO |
| **AI / Retrieval:** | Enterprise RAG, parent-child chunking, vector/full-text hybrid retrieval, RRF fusion, heuristic reranking, citation-aware context, Spring AI concepts |
| **Infrastructure:** | Linux, GitHub Actions, AWS fundamentals, CI/CD, distributed-system debugging, JVM fundamentals |

## Selected Engineering Projects

**NexusAgent – Enterprise-Aware Knowledge Assistant Backend** — Jan. 2026 – Present
*Java, Spring Boot 3.x, WebFlux, Project Reactor, PostgreSQL / PgVector, Redis, MinIO, Docker*

- Built an enterprise-aware RAG backend with MinIO raw-file storage, PostgreSQL document/chunk metadata, parent-child chunking, child-only PgVector embeddings, and idempotent ingestion / re-chunking flows.
- Implemented hybrid retrieval using PgVector vector search and PostgreSQL full-text search, then fused candidates with 1-based Reciprocal Rank Fusion (RRF), heuristic reranking, and parent-context expansion.
- Constructed citation-aware context with document IDs, filenames, parent/child chunk IDs, document-global chunk indexes, global text offsets, budget trimming, and selected-evidence preservation.
- Exposed REST/SSE query APIs and a deterministic Plan-Execute-Critique workflow reusing the retrieval/context pipeline with Reactor timeouts, retries, fallback handling, and trace IDs.
- Added Redis-backed short-lived session state, retrieval/context cache, query status, and temporary tool outputs with TTLs, tenant-aware cache keys, JSON envelopes, and graceful cache-miss/no-op fallback.
- Added enterprise-readiness slice with header-based tenant/actor context, repository-level tenant filtering, audit events, ingestion job status tracking, Actuator health/info, and trace-friendly logs.

**High-Concurrency Order & Inventory Backend** — May 2025 – Aug. 2025
*Java, Spring Boot, MySQL, Redis, message queue, RPC framework, distributed scheduler, Docker*

- Built order, inventory, and checkout services with MySQL/Redis; designed schemas and composite indexes for item lookup, order-status queries, inventory checks, and user order history access paths.
- Modeled order lifecycle transitions across pending, paid, cancelled, and expired states; added idempotency keys and duplicate-submission checks for retry-safe checkout requests.
- Implemented asynchronous expiration and compensation flows with message-queue retries and scheduled workers, handling duplicate messages and payment-vs-cancellation race conditions.
- Reduced stale-cache windows using cache-aside reads and delayed invalidation; documented Redis/MySQL consistency trade-offs and isolated slow downstream operations with custom thread pools.

**AI Agent Engineering Lab** — Oct. 2025 – Present
*Docker, Linux, Python, local / remote LLM backends, agent frameworks, GitHub Actions*

- Built a Dockerized sandbox for local/remote LLM backends, prompt workflows, and tool-use agents; compared context-window behavior, memory usage, request latency, and deployment trade-offs.
- Wrote reusable Python scripts to replay prompts and tool-call workflows, collect structured logs, and document configuration differences between local model backends and remote API providers.
- Maintained GitHub-based version control and lightweight CI checks for tests, formatting, environment validation, setup docs, and known failure modes across personal engineering experiments.

## Research Experience

**Graduate Research Assistant, RIPLE Research Group** — Riverside, CA
*Advisor: Prof. Qian Zhang, University of California, Riverside* — *Sep. 2025 – Present*

- Built Python experiment runners and model adapters for GRAphRef, a constraint-guided fuzz-testing framework for 3D mesh AI models; standardized evaluation across 8 mesh-processing systems including MeshCNN and HodgeNet.
- Ran structural mutation experiments to evaluate Valid Input Rate (VIR) and Semantic Preservation Score (SPS); generated logs, metric reports, reproducible benchmark artifacts, and LLM-output verification scripts using Benford/Zipf-style distributional checks.