

Chenyang Zhang

cyzhangv@umich.edu | +1 929-235-3841 | [Portfolio](#) | [Linkedin](#)

EDUCATION

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Master of Science, Data Information

New York, NY

Current

UNIVERSITY OF MICHIGAN, ANN ARBOR

Bachelor of Science, Double Major in Data Science and Mathematics

Ann Arbor, MI

2021-2025

Cumulative GPA: 3.8/4.0 | **Awards:** National Honor Society

TECHNICAL SKILLS

Languages: Java, Python, C/C++, TypeScript, JavaScript, SQL | **Frameworks:** Spring Boot, Flask, React, Node.js, Next.js, LangChain | **Data & Cloud:** Kafka, HBase, PySpark, AWS, GCP, PostgreSQL, Redis, HDFS | **DevOps:** Linux, Docker, Kubernetes, Git, CI/CD (CircleCI), Prometheus | **Specialties:** Microservices, REST API Design, RAG Systems, Agent/Workflow Development (LangGraph, n8n), MCP Protocol, Agile/Scrum

WORK EXPERIENCE

ALIBABA *Alibaba International – International Merchandise Team*

Software Development Engineer

Shanghai, China

Jun 2025 – August 2025

- Built a **real-time ETL pipeline** (User Insight System) with trigger-based updates to maintain dynamic profiles of **50M+ users**. Exposed profiles via **Spring Boot microservices** for multi-Agent audience segmentation.
- Opportunity Agent System: Built a Flink + ODPS Lambda-architecture ETL pipeline, establishing a 30+ core metrics library for product selection. Developed a Multi-Agent recommendation system with four specialized agents. Integrated implicit feedback loops and Prophet time-series forecasting

MEITUAN *NYSE: MPNGY*

Software Development Engineer

Shanghai, China

Mar 2025 – Jun 2025

- Developed a **supply chain optimization system** integrating image recognition (Gaussian Blur / Median Filter preprocessing) with an AI Agent pipeline. Used **LangChain** to combine RAG retrieval, behavior analytics, and domain-specific prompt engineering for procurement insights, reducing costs by 20%.
- Built a **custom React Reconciler renderer** in TypeScript (HostConfig + adapters) enabling React Native UI to run inside WeChat Mini Programs with near-zero business-logic changes. Reused state via **Redux + Redux-Saga**; optimized with static/dynamic templating, batched setData, and virtualized lists, achieving **>75% code reuse** across platforms.

UNIONPAY *World's largest card network*

Software Development Engineer

Shanghai, China

June 2024 – August 2024

- Built a **NL2SQL & Knowledge Q&A platform**: QWEN embeddings for vectorization, **multi-retrieval strategy** (BM25 + QWEN dual-tower + personalized retrieval) boosting query precision by 28%. Enhanced ranking via **Wide & Deep** modeling + Qwen1.5-32B-Chat, reducing irrelevant results by 45%.

PROJECTS

Vibehub www.vibehub.icu | Founder

2026

- Architected a **graph-based metadata store** with code-anchored entities, an **event-driven Context Bus** ingesting webhooks from Slack/GitHub/CI, a **Dynamic Rule Engine** (org → file scoping) distributing team standards to AI tools, and **MCP Server** (JSON-RPC 2.0) endpoints serving structured context.
- Built **Feature Rooms** that auto-synthesize discussions into executable AI prompts, and a **Living Documentation** system syncing branch-specific docs via **tree-sitter AST tracking** — with NLP extraction + **confidence scoring** for cross-branch **semantic conflict detection** before merge, validated with integration tests.

Nomi <https://getnomi.com/> Co-founder

2026

- A local-first AI workspace for macOS that turns the notch into an always-on entry point — one tap to capture notes, journal entries, and tasks into a personal AI brain that organizes, recalls, and searches them semantically.
- Main app pairs a coding-agent cockpit (embedded terminal, live file diffs, one-click takeover) with multi-model chat and a deep-research engine that produces cited reports ~30k LOC, 300+ tests, shipped in 3 weeks.

Caliber <https://getcaliber.dev/> Co-founder

2025

- Designed and built a full-stack platform with an LLM-powered prompt playbook extraction pipeline capturing only developer input across Claude Code, Cursor, and Codex CLI
- Shipped a multi-tenant SaaS with RBAC, custom scoring profiles, team benchmarks (signal breakdown, adoption rates, trends, task-type distribution), and that surfaces anonymized high-performing strategies for developer coaching

WhereToMeet www.wheretomeet.org Web/Full-stack

2025

- Full-stack meeting planner with real-time venue discovery via **Google Maps API**, multi-tier caching (**Redis**), user auth, event management, and voting. Optimized API performance through query optimization and caching strategies.

ADDITIONAL

Languages: Fluent in Chinese, English **Certifications:** SQL Advanced Certificate ID: C1BA102716FD