Hongyuan (Steven) Liu

Bellevue, WA liuhongyuan2001@gmail.com (647)-309-9649 liustev6.ca Yuanxyyds

EDUCATION AND SKILLS

University of Pennsylvania - Master of Engineering in AI *Aug 2025 - Current*

University of Toronto - Honours Bachelor of Science *Sep 2020 – June 2025*

- **Programs:** Double Specialist in Computer Science (AI Focus) & Data Science with High Distinction **GPA: 3.9/4.0**
- **Selected Skills:** Java, **Python**, SQL, C, C++, **Typescript**, **React**, **Node.js**, **Flutter**, FastAPI, **AWS (infra)**, **Kubernetes (CI/CD)**, Microservices, Docker, PyTorch, **Multi-Agent LLMs**, Tool Calling, RAG, LoRA Fine-Tuning, **Startups**, **Products (0 to 1)**.

PROFESSIONAL EXPERIENCE

Software Engineer in AI *Sep 2025 – Current*
Summation *Bellevue, WA, US*

- Led improvements to the **Addison Python microservices**, refactoring **multi-LLM agent** tool-calling pipelines with Pydantic AI, stabilizing SSE streaming, and redesigning chart-generation workflows, reducing latency by **50%+** and error rates by **30%+**.
- Shipped multiple **0 to 1 features** for a decision-grade AI analytics platform, including launching the **Addison Slackbot**, a production microservice handling **multimodal** inputs, generating structured AI outputs (charts, tables, code) via Slack Block Kit.
- Re-architected Summation's AG Grid data table engine, designing new cell selection and formatting models that reduced **time complexity by 80%+** and resolved **10+ edge cases**. Owned end-to-end implementation across **TypeScript FE** and **Java BE**.

Founding Full-stack Engineer *Feb 2024 – Jan 2025*
LockBox Inc. *Toronto, ON, CA*

- Co-founded, developed, and **launched** the **LockBox** on App Store and website, reaching **5,000+ downloads** and **ranking #60** in Productivity. Designed a reward-based system to reduce student phone usage by integrating iOS Screen Time APIs in **Swift** and maintaining a local-vendor reward database, resulting in an average 30% decrease in screen time.

Founder and Technical Lead *May 2023 – Sep 2024*
Campus Eats *Toronto, ON, CA*

- Founded and **led a 15+ member team** across technology, business to build **Campus Eats**, a startup that revolutionized campus dining by introducing 5+ new dining options at **30% lower costs** for students. Accepted into the **UofT Hatchery Program**, where we conducted market research with 50+ restaurants and 500+ students to develop a **business plan** and a 5-year cash flow analysis.
- Developed a robust codebase with **over 100,000 lines** across multiple products, including an **iOS/Android** app using **Flutter**, a company landing website, an admin dashboard built with **Next.js and TypeScript**, a unified backend on **Firebase**, and a similarity-based **food recommendation system** using **Autoencoder** with food embeddings on past orders.

Full-Stack Engineer Co-op *May 2023 – May 2024*
Johnson Controls *Toronto, ON, CA*

- Co-led a **two-person team** to design and deliver a high-quality, fully-functional smart home app **from scratch**, owning 20+ pages of UI/UX design in **Figma**, iOS/Android development using **Flutter**, a robust **NoSQL** database with 10+ collections in **Firestore**, and a clean-architecture backend with fully **RESTful APIs** using **Firebase Cloud Functions** in **TypeScript**.
- Co-implemented the open-source Home Assistant SDK and incorporated **Retrieval-Augmented** ChatGPT-3 API with custom tool calls to enable 100% smart device automation.

SELECTED PROJECTS

Steven Universe — Enterprise-Grade Full-Stack Personal Monorepo (50K+ LOC) *Github*

- Built and operated a **24/7 Linux-based home lab** (32 vCPU, 2 GPUs) on Proxmox, hosting an **enterprise-style microservices infrastructure** with 5+ inter-connected services across VMs and containers, managed via a unified monorepo with CI/CD.
- Designed a **GPU Service** in **Python** for 2 GPUs scheduling, idle GPU auto-allocation, and **request batching and queuing**, supporting **10+ concurrent LLM requests** via on-demand DooD worker container dispatch for training and inference workloads.
- Implemented an authenticated **File Service** using **FastAPI** as a gateway and health monitor, enforcing byte-level I/O, request-based access, and three-tier permission control over a **MinIO (S3-compatible)** storage backend.
- Built a **personal AI gateway and web server** integrating multiple services, including **StevenAI**, a fine-tuned + RAG **LLaMA 3.2** system (1,000 Q&A pairs, rank-16 LoRA), and an **LLM-powered Raspberry Pi 5 assistant** with voice interaction, custom TTS, and smart-home control via the Home Assistant API, both backed by GPU-based inference services.
- Designed and developed a **responsive, animated** **personal website** using custom **Blender 3D assets**, **Three.js**, and motion effects, serving as a unified frontend to showcase projects and integrate with backend services.