Hsiang Yu (Anna) Huang

617-319-5044 | huanganna1004@gmail.com | linkedin.com/in/hsiangyuhuang | www.hsiangyuhuang.com

Education

Boston University Boston, MA

Master of Science in Data Science *Dec. 2025*

- GPA: 3.59 / 4.0 | Relevant Courses: Deep Learning, Data Engineering, Graduate Databases, AI, Time Series
- **Award:** Winner - DS+X Hackathon 2025 (Best Overall) | Project: RhettSearch

National Taiwan University of Science and Technology Taipei, Taiwan

B.B.A. in Industrial Management & Finance (Double Major), *Minor in Computer Science* *Jun. 2023*

- GPA: 3.85/4.3 | Relevant Courses: Algorithms, Object-Oriented Programming, Machine Learning

Technical Skills

Languages & Frameworks: Python, SQL, C++, TypeScript, JavaScript, Java, R, FastAPI, Next.js, Node.js, React
AI & ML: LangGraph, LangChain, RAG, Vector Databases, PyTorch, TensorFlow, Scikit-learn, Prompt Engineering
Backend & Cloud: AWS(EC2, Lambda), Azure, PostgreSQL, MongoDB, Docker, Linux, CI/CD, Git, Redis, Kafka
Tools: PySpark, ETL, Azure Synapse, PostHog, Power BI, Looker Studio, Google Analytics, MS Clarity

Experience

AI Engineer Intern Dec. 2025 – Present

Finz | Stack: FastAPI, Faust, Kafka, CI/CD, MongoDB, AWS(EC2), Stripe, QBO *Boston, MA*

- Engineered a robust ETL pipeline to normalize raw financial data from Stripe and QBO, establishing the trusted data foundation required for accurate real-time cashflow metrics.
- Optimized NLU intent classification by refining system prompts, reducing fallback errors and improving the chatbot's ability to correctly route complex financial queries during internal testing.
- Automated deployment workflows by configuring Docker-based CI/CD pipelines in GitHub Actions, reducing manual deployment time and ensuring consistent production environments across AWS EC2.

Research Assistant – LLM Platform Sep. 2025 – Jan. 2026

BU BIT Lab | Stack: FastAPI, MongoDB, LangGraph, Prompt Engineering, OpenRouter *Boston, MA*

- Architected a context-aware "Memory System" that hybridizes short-term session history, vector-based retrieval (Top-k embedding search), and periodic conversation summarization to eliminate context drift.
- Developed an agentic intent classification system that dynamically routes user queries to specialized prompt chains (e.g., Shopping vs. General) and triggers Google SERP API for real-time product data.
- Designed a high-granularity MongoDB schema to track agent performance metrics (latency, token usage) and user interaction signals (scroll depth, clicks) for product analytics.

Full Stack Developer Sep. 2024 – May 2025

Citale (BU Spark! Launch Lab) | Next.js, SQL, Supabase, Google Maps API, Vercel, PostHog *Boston, MA*

- Developed and shipped core social features (messaging, event maps) for a beta launch, enabling the platform to support pilot collaborations with 2 local Boston businesses.
- Designed a normalized PostgreSQL schema with strict foreign key constraints to ensure data integrity for user profiles, posts, and event relationships.
- Iterated on product functionality based on PostHog analytics and direct user feedback from beta testing cycles, rapidly deploying fixes and improvements via Vercel.

Projects

Relational Database Engine Kernel | *C++20, Buffer Management, OOP*

- Implemented core storage engine components in C++20, including a Buffer Pool Manager with O(1) LRU eviction and a B+ Tree index supporting cascading splits.
- Optimized memory page layout using bitmap tracking, achieving 96% payload space efficiency by minimizing header overhead for fixed-width records.

Multi-Agent Infinite Flow Novel Generator | *LangGraph, Cloudflare Workers, Supabase*

- Designed a hierarchical agent architecture (Director, Planner, Writer) to generate long-form narratives, utilizing iterative prompt chains to drive plot progression across multiple chapters.
- Mitigated long-context drift by implementing a "Planner" agent that summarizes previous story states, allowing the system to maintain high-level narrative consistency despite context window limits.