

Ethan Liu

562-412-3749 | eliuusa@gmail.com | linkedin.com/in/ethan-liu- | github.com/WebForks | ethanliu.dev

EDUCATION

University of Pennsylvania

Masters of Science in Data Science

Philadelphia, PA

Dec. 2025 – Present

California State University Long Beach

Bachelor of Science in Computer Science

Long Beach, CA

Aug. 2021 – May 2025

EXPERIENCE

Software Engineer

June 2025 – Present

Anaheim, CA

Tone Software Corporation

- Built an MCP server backed by locally hosted LLMs to enable secure AI assistance for internal SugarCRM workflows.
- Developed a Go-based CLI client that authenticates with SugarCRM APIs and exposes CRM actions/tools to the MCP server.
- Created an internal packaging/deployment app to streamline building and shipping testable product builds for QA and stakeholder validation.
- Built a cross-platform C++ command-line utility to parse PDFs and extract embedded TrueType/OpenType font files for analysis and reuse.

Software Engineer Intern

December 2024 – June 2025

Anaheim, CA

Tone Software Corporation

- Implemented a RAG pipeline for internal documentation and code search using Qdrant vector DB and E5-Large-V2 embeddings for semantic retrieval
- Prototyped and evaluated AI coding assistants by benchmarking code completion/refactor accuracy and latency on internal tasks.
- Curated training/retrieval dataset by cleaning and de-duplicating documents, normalizing formats, and authoring prompt templates for consistent results.
- Provisioned and deployed coding/testing infrastructure across staging and production environments, improving reliability and repeatability of experiments.

Application Developer Intern

June 2023 – June 2025

Costa Mesa, CA

Siemens Digital Industries Software

- Collected and converted website documentation and PDFs into RAG-ready text with chunking + metadata to support embeddings and retrieval.
- Collaborated with a team to design the system architecture for a documentation Q&A assistant.
- Implemented an end-to-end retrieval pipeline to generate embeddings, index content in a vector database, and return top-k relevant passages to ground model responses.
- Integrated an AI assistance bot into an internal documentation website to improve interactivity and enable fast, accurate documentation lookup for users.

PROJECTS

Qwen-Image-Edit I2I LoRA

| Python, Docker, Selenium, pyTorch

- Automated image collection and preprocessing for LoRA training, producing a reproducible dataset build pipeline.
- Trained and monitored LoRA fine-tuning on a remote NVIDIA RTX 6000 GPU for 11k steps; tracked loss/outputs and tuned LR/batch settings for stability.
- Added dataset QA checks (duplicate detection, resolution thresholds, and label/caption validation) to prevent low-quality samples from entering training.

Local RAG Chatbot

| Python, LangChain, Ollama, Azure Cosmos DB (Vector Search), Docker

- Built a local Retrieval-Augmented Generation (RAG) app using LangChain with Ollama using Qwen3 model and a local embedding model to enable offline/private Q&A.
- Ingested and chunked documentation content, generated 1024-d embeddings, and stored vectors + source text in Azure Cosmos DB with vector search enabled.
- Implemented similarity search + RAG chain to retrieve top-k relevant chunks from Cosmos DB and ground responses in retrieved context; supported local development via Cosmos DB Emulator in Docker.