CHUYU YAN

647-802-4523 • c29yan@uwaterloo.ca • https://my-portfolio-oayt.vercel.app/ • https://github.com/Chuyuyan

SUMMARY

Software-focused student with hands-on experience building and deploying backend and full-stack systems, with a focus on semantic search, document processing, and automated data pipelines.

EDUCATION

University of Waterloo - 2A Electrical Engineering Graduating 2029

Relevant coursework: fundamental of programming, digital circuit,linear circuit,linear algebra, calculus 2

TECHNICAL SKILLS

Programming: C++, Python, JavaScript, TypeScript, VHDL

Backend and Systems: Flask (REST APIs), Web Automation Data Ingestion, JWT Authentication, Concurrent Programming, File Process Management

DevOps and Deployment: Git, Docker, Linux, Environment Configuration, Logging and Error Handling

Data and AI Engineering: LLM API Integration, Vector Search (FAISS), Embeddings, Semantic Search and Retrieval, Text Chunking and Indexing

Patents and Intellectual Property: Software Copyright: Automated Document Analysis Platform

PROFESSIONAL EXPERIENCE

Dingke Medical Technology - Software Engineering Intern September 2025 – December 2025

- Designed a hybrid code search pipeline combining keyword retrieval (ripgrep) and FAISS-based semantic search, returning top-6 ranked code evidence with improved precision. GitHub
- Built a scalable FAISS IndexFlatIP vector index supporting thousands to tens of thousands of code chunks (384-d embeddings), with disk persistence, incremental updates, and optional in-memory privacy mode.
- Developed a unified multi-LLM orchestration layer supporting DeepSeek, OpenAI, Anthropic, and Qwen, with response caching, automatic retries, and configurable streaming workflows for code analysis tasks.

Kangyu Medical Devices - Research and Development Assistant Intern June 2025 – April 2025

- Built a fully automated PDF metadata extraction pipeline processing 800 academic papers, extracting titles, authors, publication years, and venues into structured Excel outputs.
- Implemented robust PDF parsing and OCR workflows with image preprocessing (noise reduction and enhancement) to reliably handle heterogeneous formats and low-quality scans.
- Eliminated manual data curation, reducing processing time from 2 minutes per document to near-zero manual effort.

ACADEMIC PROJECTS

Hospital Integration Script April 2025

- Developed web automation scripts to transfer structured patient data from hospital intranet systems to a third-party case management platform.
- Automated per-patient data entry workflows involving 20–30 structured fields, processing over 800 patient records and supporting both historical case migration and ongoing intake.
- Increased data entry throughput from 30 records per day (manual) to 200 records per day via continuous automated execution.

Academic Paper Metadata Extraction Tool March 2025

- Implemented robust PDF parsing and OCR workflows with image preprocessing (noise reduction and enhancement) to reliably extract metadata from 1,000+ academic PDFs, including heterogeneous formats and partially degraded scans.
- Applied OCR selectively to 10–30% of documents where native text extraction failed, ensuring complete capture of critical fields such as title, authors, publication year, and venue.
- Generated structured Excel datasets (1,000+ rows, multi-field metadata) for downstream statistical analysis and literature review, reducing manual processing time by 40–50 hours compared to manual extraction.