Ricky Tang

437-243-5327 | rickytangdev@gmail.com | linkedin.com/in/ricky-tang-dev | github.com/rickytang666 | rickytang.dev

Education

University of Waterloo
Honours Bachelor of Software Engineering (Co-op) *Cumulative GPA: 4.0*

Technical Skills

Languages: Python, TypeScript/JavaScript, C++, SQL, Swift, C, Bash, HTML, CSS
Technologies: React, Next.js, Node.js, Express, FastAPI, Flask, RESTful APIs, LangChain, PyTorch, OpenCV, Pandas, React Native
Tools: Git, Docker, Linux, CI/CD, MySQL, PostgreSQL, Redis, SQLite, MongoDB, Firebase, Supabase, ChromaDB, GCP, Cloudflare

Experience

Hamming AI May 2026 – Aug. 2026
Software Engineer *San Francisco, CA*
- Incoming summer 2026

WAT.ai | *Python, FastAPI, LangChain, ChromaDB, TypeScript, React, DataLab* Jan. 2026 – Present
Software Engineer *Waterloo, ON*
- Developed self-correcting RAG system for **174-page EPA regulatory PDFs** in collaboration with **Bindwell (YC W25)**
- Achieved **1.00 faithfulness score** (zero hallucinations) via **DeepEval** benchmarking across golden evaluation dataset
- Engineered recursive table restoration with **100% fragment recovery** and **26% token reduction** via uniqueness filtering
- Implemented hybrid **vector + BM25** retrieval with lazy sampling, achieving **20-100x speedup** over full concatenation

WATonomous | *C++, ROS 2, Docker* Jan. 2026 – Present
Software Engineer *Waterloo, ON*
- Architected complete **ROS 2 navigation stack** including Costmap, Map Memory, A* Planner, and Pure Pursuit Controller
- Engineered **adaptive A* planner** with iterative cost relaxation, achieving **100% path success** in constrained spaces
- Amplified grid resolution by **100% (5cm)** with a **50% larger safety buffer (1.5m)**, eliminating corner-cutting collisions

Waterloo Aerial Robotics Group (WARG) | *Python, React, Flask, OpenCV, MAVLink, Docker* Oct. 2025 – Present
Software Engineer *Waterloo, ON*
- Streamlined ground station UI with one-click pause/resume for missions, eliminating manual switching for **50+ operators**
- Reduced mission failure recovery time from 30s+ to 3-5s for command pipeline operations (**85% improvement**)
- Engineered full-stack control pipeline with React frontend, Flask-SocketIO backend, and MAVLink for real-time commands
- Implemented OpenCV object detection in aerial imagery and MAVLink telemetry streaming, achieving **80%+ IoU accuracy**

Projects

Quota – VS Code Extension for Cost Optimization | *TypeScript, VS Code API, LangChain, Python, FastAPI, Next.js, MongoDB, GCP*
- Developed VS Code extension analyzing API/cloud/DB overhead with inline annotations and optimization suggestions
- Achieved <7s initial indexing and **<1s** refresh analysis via file hash caching, outperforming AI IDEs by **47x in speed**
- Implemented hybrid AST + regex parser detecting **2x more cost issues** than AI IDEs while saving **~50k tokens per check**
- Architected RAG-powered web sandbox with **LangChain + FAISS** delivering architectural recommendations in **<2s**

Nebula – AI Notes with OCR + Vector Search | *TypeScript, React Native, Python, FastAPI, Supabase, Docker, GCP*
- Developed mobile notes app with AI chat leveraging **RAG architecture** to retrieve contextual insights from user notes
- Engineered **pgvector semantic search** with **1536-dimensional embeddings** via IVFFlat indexing for **<1s retrieval**
- Implemented **Mistral OCR** pipeline extracting text and LaTeX math from images, rendering via KaTeX in markdown
- Deployed FastAPI backend on **GCP** via **Docker** with **GitHub Actions CI/CD** and **96% test coverage** using Jest & Pytest

Tark – Google Earth for Game Devs | *TypeScript, Next.js, Python, FastAPI, Redis, Prometheus, Leaflet, SciPy, Numpy, Pytest*
- Developed web app turning locations into game-ready 3D meshes in **<15 seconds** (typically weeks of manual modeling)
- Processed Mapbox elevation and satellite imagery to generate terrain meshes at **45K+ triangulated faces per second**
- Extracted **2000+ building footprints** from OpenStreetMap and generated textured .obj files for Unity/Blender workflows
- Productionized with **Redis** job store, reducing latency by **30%** via asyncio workers with **98.2% accuracy** in pytest suites

Awards

- DeltaHacks 2026: **1st Place Overall** | Hack Western 2025: **Best AI Application Built with Cloudflare**
- Hack the North 2025: **Semi-Finalist** (top 32 out of 256 teams, 1000+ hackers)