

Martin Wu

236-518-9477 | martinwu500@gmail.com | martinwu.tech | [LinkedIn](#) | [GitHub](#)

BASc Engineering student at UBC with production experience building machine learning pipelines, full-stack web applications, and embedded firmware systems across a stealth-mode AI startup and a competitive solar vehicle team.

Education

University of British Columbia

Bachelor of Applied Science in Engineering

Sep 2025 – 2029

Vancouver, BC

Technical Skills

Languages: Python, C, C++, JavaScript, TypeScript, Java, MATLAB, HTML/CSS, SQL

Frameworks & Tools: React, Flask, PyTorch, transformers, NumPy, llama.cpp, sentence-transformers, Git, GitHub Actions, Docker, InfluxDB, Grafana

ML Techniques: Large language models (LLMs), deep learning, NLP, LoRA/QLoRA fine-tuning, Q4_K_M quantization, LLM-as-a-Judge, synthetic data generation, embedding drift detection

Embedded & Hardware: STM32, SPI, CAN bus, HAL drivers, PT1000 RTD sensors, MAX31865

Developer Tools & Platforms: WSL2/Linux, STM32CubeIDE, VSCode, Vercel, HuggingFace

Experience

Co-Founder & Software Engineer

Oct 2025 – Present

Stealth-Mode AI Startup

Vancouver, BC

- Built two-stage data pipeline for local large language model fine-tuning: 8B model filters domain PDFs for relevance; 70B model synthesizes and quality-judges Q&A training pairs
- Developed LLM-as-a-Judge evaluation framework against a held-out golden anchor set, tracking factual accuracy, hallucination rate, and embedding drift with a 90% pass threshold
- Designed closed-loop retraining pipeline: model stash on quality regression, automated LoRA/QLoRA fine-tuning trigger, and human escalation with cache rollback after 3 audit failures
- Implemented VRAM-aware task scheduler with mutex locking to sequence inference, sanity checks, and training without GPU resource conflicts
- Reduced 70B model VRAM requirements from 640GB (8x A100) to 48GB (2x RTX 4090) via Q4_K_M quantization, enabling consumer-hardware deployment

Embedded Systems Engineer

Sep 2025 – Present

UBC Solar

Vancouver, BC

- Engineered PT1000 RTD temperature sensing system with MAX31865, closing a critical motor telemetry gap identified after a competition failure
- Reverse-engineered Altium PCB schematics to integrate MAX31865 hardware into the real-time driver dashboard monitoring stack
- Developed STM32 firmware in C implementing SPI communication and MAX31865 register configuration
- Achieved $<1^{\circ}\text{C}$ temperature measurement accuracy through ADC-to-temperature conversion, improving from multi-degree variance
- Integrated sensor data over CAN bus to dashboard, enabling safety monitoring independent of base station
- Configured Docker/Tailscale/InfluxDB/Grafana pipeline in WSL2/Linux via Sunlink radio for live vehicle diagnostics

Projects

Useful Tool Hub | Python Flask, JavaScript, yt-dlp, Instaloader, Vercel

2025 – Present

- Built full-stack web app with Python Flask REST API backend and JavaScript frontend; integrated yt-dlp and Instaloader for YouTube and Instagram media extraction
- Deployed backend on Vercel with CI/CD via GitHub Actions; live at martinw500.github.io/Useful-Tool-Hub

Algorithmic Trading System | Python, Pinescript, TradingView

Q1 – Q3 2025

- Developed quantitative trading strategies achieving 179% growth in backtesting with 1.05% alpha
- Implemented long/short strategies yielding 8.3% Q1 and 12.4% Q2 portfolio growth
- Built Python integration for strategy validation across 7-year historical dataset

Awards & Achievements

Mathematics Competitions (2023–2025): CSMC Top 0.5% nationally (71st, 52/60) | AIME Qualifier 2x – Top 5% on AMC 12A & 12B (118.5) | Euclid Contest Top 2% (85/100)