

Ivan Lok

ivan@ivanlok.com · +852 6432 8494 · [Personal Website](#) · [LinkedIn](#)

Professional Profile

Computer Science and General Business Management dual-degree student at the Hong Kong University of Science and Technology (HKUST), with experience spanning hardware-aware AI, computer architecture, compiler optimisation, embedded systems, robotics, and technical leadership. Experienced in performance profiling and accelerator-oriented design using TensorRT, Jetson, Gemmini, Chipyard, Chisel, LLVM, and C/C++. Led multidisciplinary robotics teams, developed reusable embedded infrastructure, managed operational budgets, and trained new members. Particularly interested in hardware-software co-design, efficient AI systems, edge AI, computer architecture, compilers, and robotics.

Education

Hong Kong University of Science and Technology (HKUST)

Dual Degree Program in Technology and Management - BEng in Computer Science and BBA in General Business Management

Sep 2023 - Expected Aug 2028

- Cumulative GPA: **4.052 / 4.3**, as of Summer 2026.
- Computer Science GPA: **4.27 / 4.3**.
- **Dean's List - every regular semester on record:** Fall 2023; Spring 2024; Fall 2024; Spring 2025; Fall 2025; Spring 2026.

Postgraduate Coursework

- **ELEC 5140 - Advanced Computer Architecture (Spring 2026): A.** Project: *Gemmini Online Attention Module and Attention Fusion*.
- **COMP 6211L - Hardware-Aware Deep Learning Computing (Spring 2026): A+.** Project: *YOLO26n TensorRT Profiling on NVIDIA Jetson Orin Nano*.
- **Selected study focus:** Hardware-Aware AI and Efficient ML Systems; Computer Architecture, Compilers, and Hardware-Software Co-Design; Embedded Systems, Edge AI, and Robotics.

Carnegie Mellon University

Undergraduate Exchange Student - Tepper School of Business

Fall 2026

- Current coursework: **10-423 Generative AI** and **10-405 Machine Learning with Large Datasets**.
- Conducting undergraduate research with CMU's **Computer Architecture and Operating Systems (CAOS) Group**.

Experience

Undergraduate Research Intern - HKUST

Automated Reasoning and Transformation of Software Research Group · Jun 2025 - Jan 2026

- Reproduced and redeployed components of an existing LLM-based compiler pass-sequence autotuning pipeline in a new computing environment, integrating the existing model workflow with LLVM pass execution and evaluation.
- Contributed to the development of a WebAssembly-to-C++ decompiler, with a primary focus on synthetic training- and test-data generation.
- Developed a diversity-driven code-generation workflow that prompted AI models to transform feature specifications and incomplete, noisy code into compilable programs containing targeted constructs, including function-pointer arrays, switch-based indirect dispatch, and pass-by-reference patterns.
- Varied feature combinations, program structures, missing components, and injected noise to generate diverse test cases targeting specific reconstruction challenges faced by the decompiler.

HKUST Robotics Team (Intelligent Racing)

Jan 2024 - Present

Senior Software Engineer

Sep 2025 - Present

- Led a multidisciplinary project team of approximately **10 members** in developing a drone and tethered ground vehicle for the National University Student Smart Car Competition, designed to respond to and cover ground beacons. As required by the competition rules, the team designed and developed the complete system in-house-including its custom circuit boards, embedded codebase, flight-control system, ground vehicle, vision stack, and ESC
- Architected the complete embedded software stack for a custom Infineon-based platform without a directly migratable codebase, and personally designed and implemented its cascaded flight-control architecture using drone dynamics and state-space modelling, integrating position, attitude, and ESC motor-speed PID control loops.
- Developed register-level libraries for **CAN bus** and **hardware I²C**, integrating them into the team's reusable hardware abstraction layer (HAL).
- Advised the succeeding Robotics Team leadership on team management, technical planning, and knowledge transfer; the succeeding team ultimately earned **one Second-Class Award and one Third-Class Award**.

Software Lead and Treasurer

Sep 2024 - Aug 2025

- Led a **9-member software team** and co-led the **24-member Intelligent Racing sub-team**, overseeing technical strategy, project roadmaps, cross-functional coordination, and a **six-figure project budget** across software, electronics, mechanics, and competition operations.
- Architected a reusable, cross-platform **C++ hardware abstraction layer (HAL)** spanning STM32, STC, and Infineon MCU platforms; unified I²C, SPI, UART, and other board-specific interfaces so sensor and control modules could be reused across robots through configuration changes rather than rewritten integrations.
- Spearheaded the first-year relaunch of the team's **National University Students Intelligent Car Race** project, setting its technical direction and guiding the team to secure **two Third-Class Honours and four Honourable Mentions**.
- Co-led a large-scale robotics training programme, developing curriculum and mentoring approximately **300 new members** in hardware and software fundamentals.

Software Engineer

Jan 2024 - Aug 2024

- Developed embedded software and interfaces for competition robotics systems.
- Contributed to software architecture, MCU integration, sensor and actuator interfaces, debugging, and on-robot testing.
- Built practical experience in C/C++, embedded development, real-time control, and multidisciplinary robotics collaboration.
- Supported competition preparation and rapid iteration under hardware and schedule constraints.
- Established the technical and organisational foundation for later promotion to Software Lead and Treasurer.

Selected Programme Leadership

Robot Design Competition 2025 - Software Department Lead

Aug 2025 - Nov 2025

- Led the full lifecycle of **all software activities** and co-led the overall Robot Design Competition programme for approximately **250 participants**.
- Orchestrated the programme from promotion and participant recruitment through delivery, including promotional campaigns, student interviews, personnel assignment, and resource allocation.
- Contributed to game-rule formulation and technical preparation while serving as a primary software lecturer and mentor throughout the competition cycle.

Robot Design Competition 2024 - Software Tutorial Lead and Lecturer

Aug 2024 - Nov 2024

- Led the complete software tutorial cycle and delivered embedded-systems lectures to approximately **100 trainees**.
- Provided hands-on technical mentorship to student groups throughout the robot design and implementation phases, supporting participants with different levels of prior programming and embedded-systems experience.
- Managed logistics and technical operations for the final competition day, supporting teams through final integration, debugging, and competition delivery.

Front-End Developer - HKUST DDP Programming Team

Sep 2024 - Jan 2025

- Led the complete redevelopment of the **React front end** for the Study Pathway Planner, redesigning the legacy interface from scratch and taking end-to-end ownership of front-end architecture, implementation, and UI/UX.
- Built an interactive drag-and-drop planning experience that allowed Technology and Management Dual Degree students to organise courses by semester and explore alternative multi-year study pathways.

Selected Technical Projects

Gemmini Online Attention Module and Attention Fusion

[Project webpage](#)

Computer Architecture / Hardware-Aware AI • Spring 2026

- Eliminated the **$O(N^2)$ DRAM round trip for attention weights** in an existing Gemmini attention pipeline, reducing attention-core latency from **611K to 245K cycles at sequence length 128 (2.49× speedup)** and accelerating the complete BERT attention sublayer by up to **1.32×**.
- Designed and implemented a custom **~700-line Chisel OnlineAttention hardware module** that performs block-wise online softmax, maintains running maximum and normalization state, approximates exponentials, reads QK^T scores directly from accumulator memory, and writes attention weights to the scratchpad.
- Extended the **Gemmini datapath and RoCC command interface** to connect the accumulator, OnlineAttention module, scratchpad, and subsequent PV computation into an on-chip attention path.
- Modified and benchmarked the **existing bare-metal C QK^T -softmax-PV workload** using Gemmini's C library and RoCC instructions, progressing from the original CPU-softmax baseline to fused and block-wise variants that remained stable through sequence length 512.
- Used **cycle-accurate Chipyard/Gemmini Verilator simulation** to validate the design and analyse why full-row fusion failed at longer sequences, evaluating block size, accumulator capacity, memory traffic, and scalar-core dispatch overhead.

YOLO26n TensorRT Deployment and Profiling

[Project webpage](#)

Edge AI / Performance Engineering • Spring 2026

- Discovered a **counter-intuitive deployment edge case**: the newer **TensorRT 10.7 was 1.44 ms slower than TensorRT 10.3** on NVIDIA Jetson Orin Nano Super in a matched FP16 compact-output pipeline.
- Used **Nsight Systems and Nsight Compute** to trace the regression to increased kernel-launch, pointwise-operation, and device-copy activity rather than uniformly slower GPU kernels.
- Designed and profiled **15 controlled YOLO26n and YOLOv8n deployment pipelines** across TensorRT 10.3/10.7 stacks, FP16/INT8 precision, and dense, engine-NMS, and compact NMS-free output paths.
- Demonstrated that **output contract and post-processing can outweigh precision gains**: an FP16 compact NMS-free pipeline achieved **7.80 ms per image**, outperforming an **8.80 ms** INT8 dense-output pipeline.
- Identified the fastest measured configuration at **7.64 ms per image at batch size 1**, with COCO mAP50-95 decreasing from **0.397 to 0.362**, and evaluated each pipeline using five-run averages over **5,000 COCO val2017 images**.

LLM-Guided Compiler Optimisation and WebAssembly-to-C++ Decompileation

LLM-Guided Compiler Pass-Sequence Optimisation

- Reproduced and redeployed components of an existing LLM-based compiler pass-sequence autotuning pipeline in a new computing environment, integrating the existing model workflow with LLVM pass execution and evaluation.
- Conducted early-stage research into extending the pipeline with additional LLVM IR and static-analysis signals, surveying candidate analyses and identifying research gaps for future pass-sequence optimisation.

Synthetic Dataset Generation for WebAssembly-to-C++ Decompileation

- Developed a diversity-driven synthetic code-generation workflow for a separate WebAssembly-to-C++ decompiler project, targeting high-level source structures lost during compilation.
- Prompted AI models to transform feature requirements and incomplete or noisy code into compilable C/C++ programs containing constructs such as function-pointer arrays, switch-based indirect dispatch, and pass-by-reference.
- Varied feature combinations, program structures, missing components, and injected noise to generate targeted reconstruction cases; exact dataset size, downstream use, and quantitative impact remain to be documented.

Cross-Platform Hardware Abstraction Layer for Competition Robots

- Architected a reusable **C++ hardware abstraction layer** spanning STM32, STC, and Infineon MCU platforms, exposing consistent interfaces over board-specific I²C, SPI, UART, and other peripheral implementations.
- Decoupled reusable sensor and control modules-including ToF ranging, optical flow, IMU, and remote-control receiver input-from platform-specific drivers, allowing the same application code to obtain data on different boards through configuration changes rather than rewritten hardware integrations.
- Enabled rapid reuse of sensor-acquisition and control software across the team's **Sumo robot, air-cushion vehicle, and drone projects**, accelerating development when projects adopted new boards or microcontroller platforms.
- Established a shared embedded-software foundation through which multiple robotics projects could integrate platform backends while preserving common higher-level sensor and control code.

Study Pathway Planner

- Redesigned and rebuilt the entire front end of an existing degree-planning platform from scratch, replacing its legacy interface with a new **React-based user experience**.
- Developed drag-and-drop semester-planning workflows that allowed students to arrange courses, revise their plans, and compare alternative pathways across multiple academic years.
- Took end-to-end ownership of front-end architecture and UI/UX, translating complex curriculum and course relationships into reusable interface components and interactive planning states.

Honours, Scholarships and Awards

Scholarships and Academic Honours

- **HSBC Global Scholarship** - 2026-2027.
- **AEON Credit Service (Asia) Scholarship** - 2025-2026.
- **Wong Chak Chui One-Belt-One-Road Enrichment Scholarship** - 2025-2026.
- **HKSAR Government Scholarship Fund - Endeavour Merit Award** - 2024-2025 and 2025-2026.
- **HKSAR Government Scholarship Fund - Talent Development Scholarship** - 2024-2025 and 2025-2026.
- **HKSAR Government Scholarship Fund - Reaching Out Award** - 2023-2024.
- **HKUST School of Engineering Admission Scholarship** - 2023-2025.
- **Dean's List, HKUST:** Fall 2023; Spring 2024; Fall 2024; Spring 2025; Fall 2025; Spring 2026 - every regular semester on record.

Competitions and Team Awards

- **Gold Medal and 11th Place - 2024 ICPC Asia Hong Kong Regional Contest.**
- **Finalist - Cathay Hackathon 2025.**
- **36th Asian Intelligent Robot Contest (2025) - Team Awards:** Co-led the HKUST Intelligent Racing sub-team, which earned **one Champion title, one First Runner-up award, and seven Honourable Mentions** across its entries.
- **20th National University Students Intelligent Car Race, South China Division (2025) - Team Awards:** Led the first-year relaunch of HKUST Intelligent Racing's programme; the team earned **two Third-Class Awards and four Merit Awards** across its entries.
- **RoboGames 2024 - Team Awards:** Member of the HKUST Robotics Team delegation, which earned **nine Gold, three Silver, and three Bronze medals** across its entries.
- **21st National University Students Intelligent Car Race (2026) - Team Awards:** Advised the succeeding HKUST Robotics Team leadership on management, technical planning, and knowledge transfer; the team earned **one Second-Class Award and one Third-Class Award.**

Technical Skills

Programming and Hardware Description

- C, C++, Python, Chisel, RISC-V assembly, Qiskit
- Familiarity with WebAssembly and compiler intermediate representations

AI, Acceleration, and Performance

- PyTorch, TensorRT, YOLO deployment, FP16/INT8 inference, model profiling
- PySpark
- NVIDIA Jetson, Nsight Compute, latency decomposition, controlled benchmarking
- Gemmini, Chipyard, hardware-aware AI, accelerator-oriented design, online attention and softmax concepts

Compilers and Software Systems

- LLVM, Clang, CMake, Git, Linux, Docker
- Compiler optimisation, program transformation, toolchain debugging, WebAssembly-to-C++ workflows

Embedded Systems and Robotics

- STM32, STC, and Infineon MCU platforms
- Embedded C/C++, HAL and driver design, peripheral integration, sensors and actuators
- Real-time control, robotics integration, debugging, FOC motor-control exposure

Engineering and Collaboration

- Technical leadership, project coordination, mentorship, teaching, budget management
- Cross-functional collaboration, conflict resolution, technical documentation, competition operations

Languages and Qualifications

- English: IELTS **8.0**, January 2026.
- Chinese: native/professional proficiency.
- Cantonese: native.