

Hwang Wei Ju (William)

Toronto, ON | william.ju@mail.utoronto.ca | +1 647-865-5552
williamhwang.com | linkedin.com/in/hwang-weiju | github.com/williamhwangweiju

Education

University of Toronto

BASc in Computer Engineering, GPA: 3.75/4.0

Toronto, ON
Expected May 2027

- **Minors:** Artificial Intelligence Engineering; Engineering Business
- **Honours:** Dean's Honour List, 8 consecutive academic terms

Experience

Machine Learning Researcher — Analog In-Memory Computing

May 2026 – Present
Toronto, ON

York University

Python, PyTorch, Hugging Face Transformers, AIHWKIT, GPT-2, IBM 3D-CIM Simulator

- Developed a sensitivity-aware placement algorithm for LLM weights across CIM tiles, reducing lifetime model-quality degradation **38%** below baseline placement; preparing a first-author manuscript.
- Restored GPT-2 accuracy under analog noise through hardware-aware fine-tuning, reducing perplexity from **3,146 to 31.5**.
- Profiled **all 49 GPT-2 projections**, discovering a **540x** noise-sensitivity spread with the LM head as the reliability bottleneck.
- Built a PCM-calibrated tile-aging simulator modeling drift, thermal variation, and faults to quantify GPT-2 degradation over hardware lifetime.

Compiler Engineer Intern

May 2025 – Apr 2026
Toronto, ON

Huawei Technologies Canada

LLVM, C++, Python, Bash, gem5, AArch64, Linux

- Developed an LLVM pass for AArch64 branch-predictor gating, cutting processor power consumption by **2.8%**.
- Designed an adaptive cache-replacement policy (TRRIP) in gem5, improving Instructions Per Cycle (IPC) by **1.65%** on production mobile workloads.
- Implemented runtime-guided instruction prefetching for mobile workloads in gem5, increasing IPC by **1.1%**.
- Automated trace and ELF generation pipelines, adopted by **20+ engineers** across the performance team.

Full-Stack Software Engineer Intern

May 2024 – Aug 2024
São Paulo, Brazil

BTG Pactual

TypeScript, Node.js, NestJS, React, AWS, REST APIs, SQL, Docker, Terraform

- Cut financial-data processing time by **40%** by automating migrations with AWS Lambda serverless pipelines.
- Shipped a full-stack banking feature enabling clients to modify payment amount, date, payer, and payee in real time.
- Built AWS SQS/SNS workflows with Terraform, decoupling payment processing and enabling independent service scaling.

Projects & Extracurriculars

University of Toronto Robotics Association — Computer Vision Engineer

Sep 2022 – May 2024

- Built a deep learning lane-segmentation model (PyTorch, YOLOv8) achieving **91.6% IoU** while reducing inference time by **64%**.
- Developed a ramp-navigation system using LiDAR, OpenCV, ROS, and waypoint planning, achieving an **85%** success rate.

Alzheimer's Detection Machine Learning Model

[GitHub](#)

- Fine-tuned an ImageNet-pretrained ResNet152 to classify **24,000 brain MRI scans** into four dementia stages, reaching **96.5% recall and F1**.
- Cut validation error by **60%** through data augmentation and hyperparameter tuning, beating a from-scratch CNN by **11 points** and a Random Forest baseline by **22**.

Geographic Information System (GIS) Software

[GitHub](#)

- Implemented C++ routing and heuristic search algorithms, reducing average route length by **14.6%**.
- Introduced multithreading to improve rendering and responsiveness, achieving a **7x** performance speedup.
- Developed an interactive GTK map interface, increasing user satisfaction from **56%** to **88%**.

Technical Skills

- **Machine Learning:** PyTorch, TensorFlow, Hugging Face Transformers, AIHWKIT, LLM inference, computer vision
- **Languages:** Python, C, C++, TypeScript, JavaScript, Java, SQL, Bash, Verilog, AArch64 Assembly, MATLAB, HTML/CSS
- **Systems & Compilers:** LLVM, compiler optimization, gem5, computer architecture, performance profiling, Linux, ROS
- **Web, Cloud & Tools:** React, Node.js, NestJS, AWS Lambda, SQS, SNS, Terraform, Docker, Git, CI/CD, Azure DevOps