Muhammad Balawal Safdar

muhammadbalawalsafdar@gmail.com | linkedin.com/in/balawal | github.com/muhammadbalawal | Montreal, QC

Education

Concordia University — Montreal, QC
Bachelor of Computer Science (BCompSc.) — *2026 - Expected 2029*

John Abbott College — Montreal, QC
DEC Computer Science — *2023 - 2026*

- **Dean's List** recipient

Technical Skills

Languages: Python, TypeScript, JavaScript, Kotlin, C#, Java, SQL, Bash
Technologies: PyTorch, TensorFlow, Next.js, Node.js, FastAPI, React Native, PostgreSQL, Docker, Kubernetes, AWS/Azure, Git, CI/CD, Linux
Spoken Languages: English (fluent), French (professional working proficiency)

Work Experience

Member of Technical Staff @ Backboard.io — 5/2026 - Present

- Built and benchmarked Backboard's **R-CLI** open-source LLM coding agent, ranked **#1 on Terminal-Bench 2.1** at **84.3%** (75/89 tasks), ahead of Codex CLI (83.4%) and Claude Code (83.1%); cut its backend's TTFT latency by **15-20%**.
- Built the **OpenAI-compatible API** layer of Backboard's PyTorch LLM inference engine: chat and completions, the Responses API, Anthropic-style messages, tool calling, structured outputs and sampling controls, so OpenAI SDK clients run against Backboard unchanged, verified end to end against the live engine.
- Improved **GPU memory efficiency** for large mixture-of-experts models on single- and multi-GPU setups, freeing VRAM for the attention KV cache without losing speed or model accuracy, which raised cache hit rates and cut time to first token on long multi-turn agent workloads.

Software Engineer Intern @ Xsolla — 8/2026 - Present

- Building a web tool that audits a published e-commerce shop in **under a minute**, scoring it across **12 checks** for speed, mobile layout, accessibility, search, and commerce visibility.
- Writing the **LLM layer** that turns audit findings into plain-language fixes, with a guard against invented findings, graded on a **40-case eval set**.

Mobile Applications Developer Intern @ Mogile Technologies (ChargeHub) — 1/2026 - 5/2026

- Shipped Android features in Kotlin and Jetpack Compose for ChargeHub, an EV-charging app with **1M+ users**, including station ratings, membership features, and map functionality in an Agile team.
- Rebuilt the **charging-session timer** to update locally every second while syncing with server data, handling multi-session edge cases, cross-device continuity, and shared ViewModel state management.
- Improved engineering productivity by stabilizing flaky **Azure Pipelines** tests, authoring custom GitHub Actions workflows, and refactoring legacy Java into idiomatic Kotlin through peer-reviewed pull requests.

CS Tutor @ John Abbott College — 3/2025 - 1/2026

- Tutored students in C# programming, focusing on OOP and data structures.
- Assisted with debugging, problem-solving, and exam preparation.

Selected Projects

cisco-pt-mcp (CUSEC 2026 Winner) — 2026

- Built an **MCP server** on top of Packet Tracer's IPC extension API (PTMP over TCP), exposing **18 tools** that let any LLM read live network state and execute configuration commands directly in the simulator.
- Published the server to PyPI as an open-source package (**2,400+ downloads**, **37 GitHub stars**), turning a manual multi-step GUI workflow into a scriptable API.

Sentinel (JACHacks 2025 Winner) — 2025

- Built a real-time traffic accident detection system using **TensorFlow COCO-SSD** and a Next.js frontend, monitoring live Quebec 511 traffic-camera feeds across **15+ regions** with bounding-box collision detection.
- Developed **full-stack architecture** integrating MongoDB (schema design and geospatial queries), RESTful APIs, Leaflet.js geographic clustering, Gemini AI incident reporting, and Twilio voice automation.

Remoto (McGill CodeJam 2025, 2nd Place) — 2025

- Built a custom **AI agent** with **12 tools** that analyzes screen state using OpenCV and Tesseract OCR to generate keyboard and mouse actions via pyautogui, enabling secure voice-controlled remote access to a machine from anywhere.
- Implemented a React frontend with speech-to-text, Supabase (PostgreSQL schema design and authentication), and **low-latency live streaming** using FFmpeg, MediaMTX, and Cloudflare.

Leadership & Community

- **Speaker** @ CUTC 2026: gave a talk on "What is Harness Engineering"; **Judge** @ CUTC Hackathon 2026
- **Volunteer** @ Hack the North 2026 (built a CTF played by **100+ students**), TechTO Hackathon, Builder Sprint: Toronto

Selected Honours and Awards

- **1st Place** @ JACHacks 2025
- **1st Place** @ MariHacks 2026
- **2nd Place** @ McGill CodeJam 2025
- **2nd Place** @ MariHacks 2025
- **3rd Place** @ SoonHacks #2
- **Best Agentic AI** @ CUSEC 2026
- **Best AI Project** @ CUHacking 2026
- **Best Design** @ MariHacks 2026
- **11labs Award** @ JACHacks 2026
- **11labs Award** @ MariHacks 2026
- **Honorable Mention** @ DawsHacks 2025