

GAUTAM KHOSLA

gautamkhosla.com | gautam.kh506@gmail.com | (+1) 437-991-2148 | github.com/GautamTalksDev |
linkedin.com/in/gautam-khosla | Ottawa, ON

SUMMARY

Computer Engineering student (Co-op, uOttawa) who builds **failure-aware systems**: real-time C++20 on QNX RTOS with no network in the safety loop, autonomous infrastructure gated by human approval, and multi-tenant SaaS built for the outage rather than the demo. Shipped production software solo for a real organization. Full write-ups, live demos, and an interactive simulation at gautamkhosla.com.

TECHNICAL SKILLS

Systems: C++20, QNX 8.0 RTOS, SCHED_FIFO real-time scheduling, POSIX message passing, CMake cross-compilation, Linux, chaos / fault-injection testing

Product: TypeScript, React, Next.js 14, Node.js/Express, FastAPI (Python), REST, WebSockets, Server-Sent Events, Tailwind

Data & Infra: PostgreSQL (Row-Level Security, Prisma), Supabase, MongoDB Atlas, Redis, Snowflake, Docker, Turborepo/pnpm monorepos, BullMQ, NATS, Cloudflare Workers, Stripe

Applied AI: Multi-agent orchestration, RAG and vector retrieval, model routing by latency/cost, evaluation and guardrails (Anthropic, Gemini, Groq, Cerebras)

SELECTED WORK

AEGIS — Edge-AI Worksite Safety on QNX RTOS (flagship)

cuHacking 7, 2026 · C++20 · QNX 8.0 · Raspberry Pi 5 · TFLite · Next.js 14 · MongoDB Atlas ·
github.com/GautamTalksDev/AEGIS

- Ran the full detection-to-alarm pipeline on-device on a real-time OS with zero network dependency in the safety loop; relay and GPIO fire at SCHED_FIFO priority 30 with voice playback deliberately subordinated at priority 8, so the alarm never waits for audio.
- Designed IPC across 4 QNX processes using trivially copyable POD message structs over MsgSend/MsgReceive, with hard ordering guarantees verified through structured latency measurement.
- Built mock event generators, load testers, and a chaos harness so the cloud gateway and dashboard shipped and were validated before hardware arrived; offline event buffer flushes in verified zero-loss chunks when the uplink returns.

Cadence — Production Inventory & Rental System, Canadian Musicians Cooperative

2026 · Built solo, requirements to deployment, for a real organization with real users · TypeScript · Next.js · Prisma · PostgreSQL · Stripe

- Took an internal tool from requirements gathering with non-technical staff to deployed production software single-handedly: role-based access control, Stripe billing, and full English/French support via next-intl.
- Made pragmatic scope decisions on what to build, defer, or solve with existing tooling; documented the system so it outlived the engagement.

HyperShift — Autonomous Cloud Orchestration Platform

2025 · Team · TypeScript · Node.js · Next.js 14 · Turborepo · BullMQ · Zod · PostgreSQL ·
github.com/GautamTalksDev/HyperShift

- Turns plain-English intent into deployed, monitored infrastructure through a 5-stage agent pipeline (Architect, Builder, Sentinel, SRE, FinOps) on durable BullMQ queues, with approval gates so nothing reaches production without a human.
- Shipped the full developer surface: CLI, workspace isolation, usage metering, per-run observability traces; shared type-safe Zod contracts across every service boundary catch integration breakage at compile time.

MetaShift — Multi-Tenant SaaS Platform, End to End

2025 · Team · Node.js · Express · React · Supabase · PostgreSQL RLS · WebSockets · Stripe ·
github.com/GautamTalksDev/MetaShift

- Built workspace-level Row-Level Security, JWT + API-key auth, and real-time conflict detection with automatic resolution streamed over WebSockets.
- Shipped the operational surface a real SaaS needs — usage metering, Stripe billing hooks, webhooks, audit logging, status page, JSON/CSV export — and owned deployment on Render/Vercel with Git-based CD.

AeroGuard — SOC Incident Triage ★ MLH Best Use of Snowflake API, HackConcordia

2025 · Python · FastAPI · React · MongoDB · Snowflake · WebSockets · github.com/GautamTalksDev/aeroguard-tech

- Dual-store architecture separating live operational reads (MongoDB) from analytical processing (Snowflake) so analytics never contend with real-time state — the design that won the award; shipped in 24h with a 7-test automated E2E suite.
- Explainable, confidence-scored triage with full audit trails, cutting operator decision time from minutes to seconds.

Chronos Cloud — Event-Driven Real-Time Platform (Technical Lead, team of 4)

uOttHack 8, 2025 · Python · NATS · MongoDB · Next.js · MapLibre · Solace PubSub+

- Architected ingestion of 14 live third-party APIs across four domains over a NATS broker, with graceful fallbacks so any upstream outage degrades the service instead of killing it; coordinated 4 engineers to a hard 36-hour deadline.

Additional: NeuroDesk — agentic middleware routing natural-language intent to specialized agents behind budget limits and approval gates (FastAPI, Redis, SSE). Omni-Context OS — stateful RAG memory across Slack, email, and web; Top 9, Backboard.io × McHacks.

EXPERIENCE & LEADERSHIP

Data Processor & Software Developer — Canadian Musicians Cooperative

Jul - Aug 2026

- Built Cadence (above) solo; audited and repaired data-integrity failures across cooperative tooling; automated recurring workflows with Apps Script, Slack webhooks, and scheduled triggers.

Financial Sales Representative — Kognitive Sales Solutions × Canadian Tire

Oct 2025 - Jan 2026

- Explained regulated financial products to high-volume customers under strict compliance — daily practice making complicated things understandable to people who did not ask for the details.

Class Representative, GNG1105 — uOttawa Engineering Students' Society

Sep 2024 - May 2025

- Primary liaison between students and faculty; gathered feedback, identified recurring issues, escalated through departmental meetings until resolved.

EDUCATION

B.A.Sc. Computer Engineering (Co-op) — University of Ottawa

Sep 2024 - Present · Ottawa, ON

- Data structures & algorithms, databases, software engineering, computer architecture, discrete mathematics, probability & statistics.