Sanjay Subramanian

☎ 416-417-4309 | sanjaysubb2006@gmail.com | linkedin.com/in/sanjay-subramanian-bb2615315/ | github.com/Sanjay617

Education

University of Waterloo — Waterloo, ON
Bachelor of Mathematics and AI, Minor in Computer Science — *Expected Graduation: April 2028*

Technical Skills

Languages: Python, Go, C, C++, SQL (Postgres, MySQL), JavaScript, TypeScript, HTML/CSS
Frameworks: React, Next.js, Node.js, Flask, FastAPI, GraphQL, Ant Design
Developer Tools: Git, Docker, Kubernetes, AWS, Terraform, Redis, GitHub Actions, VS Code
Libraries: pandas, NumPy, Matplotlib, LangGraph, crewai, Jest, React Testing Library

Experience

Software Developer — Aug 2026 – Present
Macaca AI — *Remote*

- Built backend features across a 6-service multi-tenant AI agent platform serving WhatsApp, email, and web chat.
- Designed async job pipelines on Redis Streams with retry, backoff, and dead-letter queues to isolate slow providers.
- Added Go REST endpoints and schema migrations behind a single-owner Postgres gateway, keeping 5 workers driver-free.
- Extended a RAG pipeline that chunks and embeds uploaded documents into Qdrant for per-tenant knowledge retrieval.

Web Developer, Software Team — May 2026 – Present
TEDxUW — *Waterloo, ON*

- Building the TEDxUW 2026 website in Next.js and React, covering the home, speaker, sponsor, and ticketing pages.
- Optimized media-heavy pages with Next.js image optimization and lazy-loaded video for fast, responsive mobile loads.
- Developing the event ticketing flow and internal tools used by Logistics and Corporate Relations to manage the event.
- Partner with Marketing and Speaker Relations to publish speaker profiles, past talks, and sponsor content each cycle.

Software Engineering Intern — Jan 2026 – Apr 2026
U Plus — *Richmond Hill, ON*

- Rebuilt the merchant onboarding flow in React/TypeScript, lifting sign-up completion by 11% and cutting setup time by 35%.
- Replaced sequential DB calls with batched PostgreSQL queries on 3 FastAPI endpoints, reducing p95 latency by 18%.
- Right-sized EC2 instances and added request-driven auto-scaling policies, cutting monthly cloud costs by $3,800.
- Integrated OpenTelemetry tracing into the backend, reducing average time to root-cause production incidents by 15%.

Software Engineering Intern — Jan 2025 – Apr 2025
Terre Sky — *Remote*

- Built a GitHub Actions CI/CD pipeline with automated rollback, reducing new feature deployment time by 14%.
- Wrote modular Terraform scripts to provision all AWS infrastructure, cutting staging setup time from 1 hour to 5 minutes.
- Engineered a Rust data ingestion service for a logging platform, processing 5,000+ events/second at sub-10ms latency.
- Containerized 6 microservices with Docker and published images to ECR, enabling consistent dev/staging/prod deployments.

Projects

Model-Based RL for Humanoid Locomotion | *C++20, MuJoCo, CMake, GoogleTest, Python (PPO baseline)* Sep 2026 – Ongoing

- Wrote a C++20 neural network engine from scratch with own matrix math and backprop, verified by gradient checks.
- Trained a 5-model ensemble world model on cart-pole and planned with it using a multithreaded CEM planner.
- Training a 12-motor humanoid walking controller inside the learned model, targeting a 10 m walk in MuJoCo simulation.
- Benchmarking sample efficiency against a PPO baseline over 5 seeds with confidence intervals.

Expense Tracker & Budget Analytics Dashboard | *React, TypeScript, FastAPI, PostgreSQL, Docker* September 2025

- Built full CRUD with AG Grid editable tables and real-time validation, reducing manual data entry errors by 25%.
- Implemented JWT auth with bcrypt hashing for 100+ accounts with row-level data isolation.
- Delivered spend analytics and budget tracking, improving user budget adherence by 30% within the first month.
- Containerized the stack with Docker Compose, cutting new-contributor setup from manual steps to a single command.

Certifications

AWS Certified Cloud Practitioner **AWS AI Practitioner** **(OCP) – Java SE 11**