

# SACHIN VED GUPTA

☎ 437-217-4947 ✉ [SachinVedGupta@gmail.com](mailto:SachinVedGupta@gmail.com) in [linkedin.com/in/Sachin-Ved-Gupta](https://www.linkedin.com/in/Sachin-Ved-Gupta) 🐙 [github.com/SachinVedGupta](https://github.com/SachinVedGupta)  
📄 [medium.com/@SachinVGupta9](https://medium.com/@SachinVGupta9) 👤 [SachinVedGupta.com](https://SachinVedGupta.com)

## Education

### McMaster University

September 2023 - April 2027

Bachelor of Software Engineering Co-op

Hamilton, Canada

- GPA: 3.9/4.0, **Provost's Honour Roll Medal** (perfect 4.0 GPA in 1st year), **\$10000** Scholarship (OpenText, Rogers)
- Relevant Courses: Data Structures & Algorithms, OOP, Software Design, Computer Architecture, Linear Algebra

## Experience

### Machine Learning Intern

September 2025 – Present

Shopify

Toronto, Canada

- Incoming Fall 2025 on the Shopify Magic Generative AI team

### Machine Learning Team Lead

January 2025 – Present

McMaster Aerial Robotics & Drone Club 📄

Hamilton, Canada

- **Fine-tuned YOLOv11** on a custom balloon/bucket dataset to better identify fire markers (+2.6% mAP, +4.4% recall)
- Architected a camera-to-NVIDIA Jetson Nano-to-flight controller pipeline and optimized it by deploying a TensorRT object detection model with **FP8 quantization**, achieving real-time, **low-latency inference with a 3.2× speedup**
- Led a 20-member team across autonomous drone development and Gazebo simulation testing to prepare for SUAS 2026

### Machine Learning Intern

April 2025 – August 2025

Nokia Bell Labs

Ottawa, Canada

- Built a Retrieval-Augmented Generation (RAG) system using LangChain and Pinecone embeddings to leverage historical errors and solutions from JIRA for generating context-aware plans of action to solve new issues (adopted by 8+ teams)
- Enhanced the system with **LangGraph-based AI agents** that connected to the dashboard, identified failing VMs, and autonomously executed recovery actions (primarily Linux commands), automatically resolving over 60% of daily issues
- Analyzed industry research papers on Generative AI's impact on 5G traffic and **leveraged Pandas, NumPy, and Matplotlib** to visualize trends, revealing a strong cross-correlation ( $r = 0.72$ ) between AI demand and 5G network load
- Developed an **RNN-based predictive model (4.4% MAPE)** using TensorFlow to forecast AI-driven 5G network traffic, enabling Nokia to make data-driven decisions that led to a significant increase in infrastructure investment

### Software Engineering Intern

April 2024 – August 2024

Nokia

Ottawa, Canada

- Automated OpenStack VM provisioning using modular Python scripts, resulting in a **98.3% reduction** in manual work
- Developed a Chrome extension to automate log parsing and identify root causes of regression test failures, reducing average bug resolution time by **64%** and improving team debugging throughput through dynamic error highlighting

## Impactful Projects

### AWS - Llama LLM Inference Optimization Contest 📄 | AWS (Neuron Kernel Interface, Trainium EC2), Pytorch, Numpy

- Optimized performance by re-implementing fundamental transformer inferencing operations like Matrix Multiplication and Scaled Dot Product Attention to **utilize NKI** which is more efficient for the Trainium EC2 chips it's running on
- Utilized the benchmarking tool to monitor performance metrics (accuracy, NKI flops, throughput) - **now 1.13x faster**

### LearnBridge - AI Agent for Students 📄 | Google Agent Development Kit, GCP, Python, Google Classroom & Calendar API

- Architected **ReAct-based multi-agent workflows** with memory, sequential/parallel execution, tool integration (Google Classroom & Calendar), and prompt tuning for intelligent, context-aware task handling in a Streamlit UI
- Enabled faster assignment starts by extracting key concepts from assignments/announcements/slides, returning formatted outputs (graphs, tables, bolded summaries) (80% faster content access), and auto-adding deadlines to Google Calendar

## Skills

**Programming:** Python, C++, SQL, JavaScript, Linux, Docker, PostgreSQL, CypherQL (Neo4J), Google ADK, LangChain

**Libraries:** PyTorch, Tensorflow, Triton, CUDA, Flask, Pandas, Numpy, Matplotlib, Scikit-Learn, OpenCV, Hugging Face

**Certifications:** AWS Certified AI Practitioner, AWS Certified Cloud Practitioner, Nokia Bell Labs - Data Driven 5G 📄

**Highlights:** Won HackMIT (1000 competitors) 📄, Ignition Hacks Speaker 📄, AI articles 📄, Attended ASPLOS & ICSE

**Google Developer SC:** Leading a 10-member team across agile development of an open-source AI Browser Agent 📄