

Raghav Yadav

📞 647-892-4001 | ✉️ raghav.yadav@mail.utoronto.ca | 📁 Portfolio | 🔗 LinkedIn | 🐙 GitHub

EDUCATION

University Of Toronto

Bachelor of Applied Science and Engineering, Computer Engineering

Toronto, ON

Sep. 2023 – May 2028

EXPERIENCE

Google Summer of Code – Intel (OpenVINO Toolkit)

May 2025 – August 2025

Machine Learning Open-Source Engineer

Toronto, ON

- * Built and deployed **RESTful + gRPC** embedding endpoints in **OpenVINO Model Server** for **multimodal** inputs.
- * Integrated preprocessing with **OpenVINO runtime**, reducing multimodal pipeline latency by **45%** and scaling throughput by **3×**.
- * Optimized inference on **Intel iGPUs**, reducing latency from **3600ms → 55ms (65× speedup)**.
- * Containerized services with **Docker** and built **CI/CD pipelines**; enabled **cross-modal retrieval** using **HuggingFace + LangChain**.

Founding Engineer – Stealth AI Startup

May 2025 – Present

Stealth Venture

Toronto, ON

- * **Leading development** at an early-stage stealth venture focused on building a **context-aware personal AI assistant** designed for daily human interaction.
- * **Architected a multimodal system** leveraging **large language models**, **vector databases**, and **real-time contextual signals**.
- * **Secured \$5,000 non-dilutive microgrant funding** to support prototyping and initial research.

Software Engineering Intern

May 2024 – August 2024

University of Toronto Entrepreneurship Hatchery

Toronto, ON

- * Engineered a high-throughput RAG system over **100K+ documents** using LangChain, Pinecone, and FastAPI
- * Reduced API latency from **1100ms to 180ms** using streaming and async execution
- * Deployed scalable services with Docker and CI/CD, enabling **zero-downtime rollouts**
- * Built monitoring and logging pipelines to support **20K+ monthly queries** in production

PROJECTS

ResumeAI | Python, Next.js 14, TypeScript, spaCy, scikit-learn

GitHub

- * Built an AI-powered tool with NLP ML to deliver ATS checks, keyword relevance scoring, and FAANG-optimized resume feedback.

ShellPilot | Python, Javascript, Typer, Linux CLI, AI/LLM APIs

GitHub

- * **Built** a CLI-based assistant enabling natural language execution of Linux commands with **DeepSeek R1** integration.
- * **In Progress:** Expanding support for additional **LLM providers** (OpenAI, Anthropic, Ollama) and web frontend integration.

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, C++, C, SQL, Bash Script, HTML/CSS, Java

AI/ML: PyTorch, HuggingFace, scikit-learn, Keras, OpenCV, NumPy, SciPy, Pandas, Matplotlib, Seaborn, TensorRT

Frameworks: React, NextJS, Node.js, Express.js, Flask, FastAPI, TailwindCSS

Utilities: Git/GitHub, Amazon Web Services (AWS), Google Cloud Platform (GCP), CI/CD, MongoDB, PostgreSQL, Spring Boot