

Vividh Mahajan

548-922-2600 | v7mahaja@uwaterloo.ca | [linkedin.com/in/vividhm](https://www.linkedin.com/in/vividhm) | github.com/Lasdw6 | [Portfolio](#)

TECHNICAL SKILLS

Languages: Python, C++, Typescript, HTML, CSS, SQL, Git

Developer Tools: Docker, Terraform, Azure, Google Cloud, Pinecone, LinuxCLI, GitHub, REST APIs

Libraries & Frameworks: Pytorch, FastAPI, Django, Langchain, Huggingface, Numpy, React.js, Next.js, MongoDB

EDUCATION

University of Waterloo

Waterloo, ON

Bachelor of Mathematics in Combinatorics and Optimization, minor in Computer Science

Sep. 2023 – Present

Presidents Scholarship Recipient, Computer Science Club Syscom Engineer, Tech+ Club Backend Engineer

Relevant Coursework:

Tools for Software Development, Object-Oriented Software Development, Linear Programming, Optimization

External Courses: Deep Learning Specialization - Deeplearning.ai (**125 hours** [↗](#))

EXPERIENCE

Software Engineering Intern

Sept. 2025 – Dec. 2025

Manulife

Toronto, Canada

- Developing an internal agent in Python + FastAPI to automate incident tracking for 800+ employees, reducing manual processing time from **2 hours to 10 minutes**
- Building and scaling infrastructure with Terraform on Azure to support **10k+ daily incident management requests**

Software Engineering Intern

May 2025 – Aug. 2025

GOQii

Mumbai, India

- Developed a medical RAG assistant using LangChain + Pinecone to query 50k+ medical records, reducing doctor lookup time from **5 mins to 30 seconds**
- Redesigned unstructured data ingestion pipeline using Python + regex + schema validation, **cutting parsing errors by 40%**

Machine Learning Engineer Intern

Dec. 2024 – Feb. 2025

The Innovation Story

Remote, Canada

- Designed a lightweight, graph-based recommendation algorithm for deployment in a mobile education app
- Trained a YOLOv11 model using PyTorch on a custom PCB component dataset (350 images across 7 classes), **achieving 94.3% mAP@0.5 with ~206 ms CPU inference**, reducing lab-setup time by 67%

Software Engineering Intern

Sep. 2024 – Dec. 2024

Electron Online

Mumbai, India

- Worked on an Analytical Marketing Platform to process **10,000+ social media posts** monthly to generate reports, helping businesses understand customer sentiment on their products, using **Django** and **React**
- Built an end-to-end data pipeline to collect, process, and store over **100GB of social media data/week** using web scraping and API integrations.
- Optimized **Natural Language Processing (NLP)** model accuracy by **15%**, using LLM prompt engineering and fine-tuning, for sentiment analysis

PROJECTS

Housing Search Agent | *Python, Playwright, FastAPI, DuckDuckGo*

Jan. 2025 – Present

- Built a browser-based AI agent using **Playwright** to automatically find rental listings (e.g., Rentals, Kijiji, Roomies), based on user location and budget preferences
- Integrated **LLM-driven classification** with **RAG (Supabase)**

Tea Tree Chat [\[askteatree.chat ↗\]](#) | *FastAPI, Next.js, Postgres, REST APIs*

June 2025 - Present

- Built a full-stack AI chat app supporting GPT, Claude, and Gemini via OpenRouter BYOK
- Reduced LLM response latency by over 60%** by implementing backend response caching, significantly improving perceived app performance
- Shipped a responsive, minimal UI with dynamic LLM switching using Next.js and deployed the app end-to-end