

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, AWS, Pinecone, LinuxCLI, GitHub, REST APIs

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

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

Coursework: 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 - Enterprise Technology, GenAI

Toronto, Canada

- Owned backend orchestration and reliability for a production AI-driven incident triage system used by 1,000+ employees and **reducing average triage time from 2 hours to 5 minutes**
- Re-architected backend infrastructure after identifying scaling bottlenecks, introducing Redis caching, connection pooling, and Terraform-managed infrastructure to reliably support **10k+ daily requests**
- On-call during production deployments, diagnosing live failures and driving the design and implementation of an observability layer (latency metrics, structured logging, execution metadata), improving agent debugging

Tech: Python (FastAPI), React.js, Azure, Redis, Terraform, LLM APIs

Software Engineering Intern

May 2025 – Aug. 2025

GOQii

Mumbai, India

- Owned the design and development of a medical assistant enabling doctors to query 50k+ patient records, reducing information lookup time from **5 minutes to 30 seconds**
- Redesigned the unstructured data ingestion pipeline for medical reports, reducing parsing errors by **40%** and improving data reliability

Tech: Python, RAG, Vector Databases, Data processing, AWS

Machine Learning Engineer Intern

Dec. 2024 – Feb. 2025

The Innovation Story

Remote, Canada

- Designed and implemented a lightweight, graph-based recommendation algorithm optimized for deployment in a mobile education application.
- Owned model development and evaluation for a custom computer vision system, **achieving 94.3% mAP@0.5 with 206 ms CPU** inference and reducing lab setup time by **67%**.

Tech: PyTorch, Computer Vision, Graph Algorithms

Software Engineering Intern

Sep. 2024 – Dec. 2024

Electron Online

Mumbai, India

- Owned backend development for an analytics platform processing **10,000+ social media posts per month**, enabling automated sentiment and performance reporting for client businesses.
- Designed and built a large-scale data ingestion and processing pipeline handling **100GB+ of data per week**, improving sentiment analysis accuracy by **15%** and increasing reliability of downstream reports.

Tech: Python, React.js, Google Cloud, LLM APIs

PROJECTS

Agentic Personal Assistant | *Python, Langchain, Pinecone, AWS, Docker, Google APIs*

Jan. 2025 – Present

- Implemented a stateful personal agent that consumes contextual signals (email, calendar, location) to infer user intent and proactively trigger actions, such as detecting missed commitments and proposing reschedules.
- Iterated on the agent's memory and decision flow to combine long-term context with real-time inputs, reducing false positives and improving reliability of proactive suggestions.

Tea Tree Chat [[askteatree.chat](#) [↗](#)] | *Python, Next.js, Postgres, Openrouter, Clerk*

June 2025 – Present

- Shipped a full-stack multi-LLM chat application with a BYOK model and improved perceived response latency by **60%** through backend request caching and optimization.