

Umair Arham

Toronto, ON | 647-739-2107 | contact@uarham.me | linkedin.com/in/umairarham | github.com/umairx25

EXPERIENCE

CBRE

June 2026 – August 2026

AI Engineer Intern | Python, LangGraph, Google Cloud, PostgreSQL

- Developed an AI automation engine using **Python** and LangGraph, saving **20+** hours of developer time/week
- Transformed **100+** documents into structured business logic using a **LangGraph** based multi-agent system
- Cut **90%** of incorrect AI logic reaching production by integrating **confidence-based** human approval steps

UrbanFood Alliance

May 2025 – August 2025

ML Engineer Intern | Python, LangChain, Azure, Docker, React

- Built **Langchain** AI agent to optimize onboarding for **30+** users by surfacing company info and relevant resources
- Integrated **AI agent** with company knowledge using a LlamaIndex **RAG** pipeline across **50+** internal documents
- Cut information lookup time by **30%** by integrating tools connecting to external APIs and databases
- Scaled the application to support **100+** users by containerizing with **Docker** and deploying on **Azure**

Department of CS, University of Toronto

September 2025 – December 2025

Software Engineer | Typescript, Node.js, Express.js, PostgreSQL

- Deployed a course platform to enable **200+** students to coordinate coursework by containerizing app with **Docker**
- Cut access errors across **20+** workflows using automated tests on isolated **PostgreSQL** databases
- Prevented **5+** faulty pull requests from merging by enforcing automated tests through **GitHub Actions**

EDUCATION

University of Toronto

Bachelor of Science in Computer Science

Expected May 2027

PROJECTS

Invisible - Secret Management Platform | Python, Go, Google Cloud, PostgreSQL, Redis

- Cut secret retrieval latency to **<2s** by developing a **Go-based CLI** and Python/Node SDKs for on-demand access
- Developed secrets management platform used by **10+ teams** to securely store and access production credentials
- Reduced repeated **interactive logins** by implementing long-lived **JWT-based** device authentication
- Deployed stateless **REST API** on **Google Cloud Run** backed by **PostgreSQL** for multi-tenant secret access

NotedAI | AWS S3, AWS SQS, AWS RDS, pgvector, Docker

- Developed cloud-native meeting processing pipeline backed by **AWS**, processing **40+** min recordings
- Architected an event-driven pipeline using **S3** and **SQS**, supporting **5+** concurrent uploads via async processing
- Integrated vector-based semantic search using **pgvector**, enabling contextual retrieval across meeting transcripts

Tabi - Agentic AI Extension | Python, Langchain, Gemini, FastAPI, Chrome API

- Achieved **60+** active users and **100+** daily interactions by publishing AI extension to the Chrome Web Store
- Built real-time tab categorization with **<1s** latency by connecting extension to ML model via **FastAPI** backend
- Reduced LLM output errors by **35%** by enforcing schema-constrained generation with **Pydantic** and **Gemini**
- Achieved **90%** classification accuracy across **10+** categories by fine tuning an ML model with **sci-kit learn**

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, Java, SQL, C

Developer Tools: Git, GitHub Actions, Docker, Linux, Bash

Frameworks: FastAPI, Django, Express.js, Node.js, React, React Native, SQLAlchemy

Cloud & Databases: Google Cloud Platform, AWS (S3, SQS, RDS), PostgreSQL, MySQL, MongoDB, Redis

AI & ML: LLMs, Embeddings, LangChain, LangGraph, OpenAI API, pgvector, Claude Code, Codex, Scikit-learn

Concepts: Distributed Systems, Microservices, Async Processing, Caching, RESTful API, Object-Oriented Programming, Data Structures and Algorithms