

Aaryan V Bhalla

+1 (942) 444 1412 | avbhalla@uwaterloo.ca | bhalla.info | github.com/aaryanved | in/aaryanved

EDUCATION

University of Waterloo

Bachelor of Computer Science (Honours, Co-Op)

Sep 2025 – Present

Waterloo, ON, CA

SKILLS

AI & ML: Python, LangChain, LangGraph, RAG, OpenAI API, TensorFlow, Transformers (HuggingFace), FAISS, scikit-learn, Pandas, NumPy, Pydantic

Backend & Databases: FastAPI, Flask, Node.js, PostgreSQL (asyncpg, pgvector), SQLite, REST APIs, SSE, asyncio

DevOps & Cloud: AWS, Git, CI/CD, Docker, GitHub Actions, Bash, Linux

Frontend: React, Angular, TypeScript, JavaScript, Tailwind CSS, HTML, CSS, Chart.js

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

EXPERIENCE

AI Engineer Co-op

Decimal Technologies

May 2026 – Present

Delhi, India

- Designed and built a production-grade LLM backend using FastAPI, LangGraph, LangChain, and OpenAI APIs, streaming real-time AI responses to an Angular frontend via Server-Sent Events
- Architected a modular RAG agent pipeline with intent-based routing (analytics, FAQ, lead detail, clarification), secure parameterized SQL generation, and pgvector semantic retrieval on a structured PostgreSQL schema
- Implemented end-to-end auth (sign-up, session management, RBAC) and full database layer with schema migrations, conversation history, and per-turn telemetry tracking token usage, latency, cache hits, and cost

Co-Founder and Event Director

SHISTECH

Apr 2021 – Apr 2025

Delhi, India

- Co-founded and scaled a technical organization to 50+ members, building all internal software tools, automation systems, and data pipelines from scratch to support a 6,000+ participant annual event
- Designed and shipped a full-stack cryptic hunt platform using Node.js, EJS, and vanilla JavaScript, sustaining 100+ concurrent users across a live competitive event with zero downtime
- Drove end-to-end technical ownership across software, data, and operations infrastructure for four years - building, iterating, and maintaining systems with real users and real stakes

PROJECTS

Pulse AI | FastAPI, LangGraph, PostgreSQL, asyncpg, pgvector, OpenAI API, SSE

May 2026

- Built an LLM-powered banking analytics chatbot backend, replacing a mocked system with a fully functional agent capable of answering complex queries on CASA lead data via natural language
- Implemented intent-based routing across agent nodes (analytics, FAQ, lead detail, clarification), dynamic SQL generation with injection-safe parameterization, and selective retrieval using pgvector embeddings
- Designed per-turn observability: token usage, latency, cache hit tracking, and cost telemetry persisted to PostgreSQL alongside full conversation history and role-based access enforcement

The Originality Engine | Python, NLP, Transformers, FAISS, Pandas, NumPy

Feb 2025

- Built a research-grade ETL and ML pipeline to ingest, embed, and score arXiv abstracts using transformer sentence models and FAISS nearest-neighbour indexing
- Designed a modular, script-driven architecture with clearly separated ingestion, transformation, indexing, and scoring stages; evaluated outputs against ground-truth similarity metrics

Grounded | JavaScript, Gemini 2.5 Flash, ElevenLabs, face-api.js, Web Audio API

Mar 2026

- Shipped a real-time multi-service AI pipeline at Hack Canada 2026: webcam input, ML-based stress detection, LLM response generation, and TTS voice output, orchestrated end-to-end in under 24 hours
- Engineered async coordination across three external APIs (Gemini, ElevenLabs, face-api.js) with concurrency guards, fallback logic, and graceful error recovery that held up reliably through a live demo

Disease Detection | Python, TensorFlow, Flask, Pandas, scikit-learn, Raspberry Pi

Aug 2023

- Trained a dual-mode TensorFlow diagnostic model applying image preprocessing, normalization, augmentation, and iterative hyperparameter tuning to improve generalization across test sets
- Deployed the model end-to-end on Raspberry Pi via a Flask API, implementing confidence thresholding, structured JSON responses, and full inference pipeline management for real-time portable use

ACHIEVEMENTS

National Winner - IIT Delhi Youth Ideathon

2024

Placed first nationally at the Indian Institute of Technology Delhi's Youth Ideathon, competing against university teams across India in a technology and innovation challenge.