

# Yash Gupta

Calgary, AB | guptayash.connect@gmail.com | 1-807-356-4074 | yashcodes.dev | linkedin/yash-gupta

## Technical Skills

---

**Languages:** Python, SQL, Scala, TypeScript, C++, Bash  
**AI/ML:** LLMs, RAG, Graph-RAG, Multi-Agent Systems, Prompt Engineering, LangChain, HuggingFace, PyTorch, TensorFlow  
**Cloud & Infrastructure:** AWS (Lambda, Step Functions, SageMaker, Bedrock, ECR), GCP (Cloud Run, Vertex AI), Docker, CI/CD, Serverless Architecture  
**Data & Systems:** Pinecone, Neo4j, OAuth 2.0, FHIR APIs, REST APIs, WebRTC

## Experience

---

**AI Engineer (Product), Monark** May 2025 – February 2026

- Led end-to-end design and deployment of a **production LLM-powered voice AI roleplay system**, re-architecting the pipeline to increase **transcript retention by 50%**.
- Owned hallucination reduction strategy, decreasing **model errors by 95%** using **structured prompt guardrails** and an automated **LLM-as-a-judge** evaluation framework.
- Architected and led implementation of **scalable serverless AI infrastructure** on AWS, achieving **sub-60s** latency and enabling fully automated **CI/CD** pipelines.
- Designed and deployed a **multi-agent RAG system** leveraging **Vertex AI Vector Search** for structured conversational querying across meeting datasets.
- Directed migration of AI workloads to **GCP Cloud Run**, **reducing infrastructure costs by 60%** while maintaining **90% F1** in PII redaction and speaker identification pipelines.

**AI Researcher, Aurora Constellations** May 2023 – September 2024

- Led research and development of **Graph-RAG and knowledge graph systems** to generate **structured clinical treatment plans** from **large-scale MIMIC** datasets.
- Defined research direction and system architecture for **scalable RAG services** integrated into **Scala-based clinical DSL** tooling.
- Directed secure integration of **OAuth 2.0** and **FHIR** pipelines with **Epic EHR** systems for **HIPAA-aligned** clinical data ingestion.
- Designed and trained **Graph Neural Networks** on **heterogeneous clinical data**, achieving **AUROC 0.74** for ICU mortality prediction.
- Translated **oncology treatment protocols** into **interactive AI-driven decision-tree systems** to reduce **clinician cognitive load**.

**Software Engineer, Aurora Constellations** May 2021 – May 2023

- Built and maintained **CI/CD pipelines** and **structured logging** for **Scala Play** backend services to improve reliability and observability.
- Developed **secure authentication tooling**, including automated **JWT certificate provisioning** for service-to-service and user auth flows.
- Implemented a **wake-word voice assistant** (**Picovoice + Google Speech-to-Text**) to enable hands-free clinical workflow automation.

## Education

---

**Lakehead University**, MSc in Computer Science (AI Specialization) Thunder Bay, ON  
• Vector Institute Scholarship for Artificial Intelligence

**Lakehead University**, BEng in Software Engineering Thunder Bay, ON  
• Dean's Scholar Award, President's List