

ASHOK TAK

GenAI Engineer | Multi-Agent Systems | MLOps

ashoktak95@gmail.com | 647-913-2127 | Toronto, ON, Canada | [linkedin.com/in/ashoktak](https://www.linkedin.com/in/ashoktak)

PROFESSIONAL SUMMARY

GenAI Engineer with **8+ years** of experience in Data and AI — including **2+ years** specializing in production GenAI systems: **multi-agent architectures, RAG pipelines, LLM-powered applications, and prompt engineering frameworks** at enterprise scale. Proven ability to architect and ship end-to-end GenAI solutions: from agentic workflows (LangChain, LangGraph) and cloud AI services (AWS Bedrock, SageMaker) to containerized API deployments (FastAPI, Docker) and automated CI/CD (GitHub Actions). Strong MLOps foundation — experiment tracking, model lifecycle management, drift monitoring, and production-grade reliability. **Databricks Certified ML Associate & Data Engineer Professional**. Translates complex business requirements into robust, scalable AI engineering. MAsC, University of Toronto. 4 peer-reviewed publications.

CORE TECHNICAL SKILLS

GenAI & LLMs	LangChain, LangGraph, LlamaIndex, multi-agent systems, RAG architectures, prompt engineering, Hugging Face Transformers, sentence-transformers, vector embeddings, pgVector, AWS Bedrock (Agents, Knowledge Bases, Guardrails), OpenAI APIs
APIs & Deployment	FastAPI, REST APIs, Docker, containerization, Kubernetes, microservices design, API versioning, production SLAs, batch & real-time inference
MLOps & CI/CD	MLflow, Model Registry, Experiment Tracking, model versioning, drift monitoring, automated retraining, CI/CD for ML (GitHub Actions, GitLab CI), Terraform, DevSecOps
Cloud & Infra	AWS (Bedrock, SageMaker, S3, Glue, EMR, EC2), Azure (ADF, Synapse, AKS), GCP (BigQuery, Vertex AI, Dataproc), Databricks
ML Frameworks	scikit-learn, PyTorch, TensorFlow, Keras, XGBoost, LightGBM, LSTM, ensemble methods, SMOTE, NLP, time-series modelling, classification, clustering
Data & Pipelines	PySpark, Apache Spark, dbt, Apache Airflow, Kafka, Delta Lake, Great Expectations, feature stores, data contracts
Programming	Python (clean architecture, packaging, testing, performance tuning), SQL (advanced), PySpark, Bash, REST APIs
Responsible AI	bias/fairness analysis, model explainability, Bedrock Guardrails, content safety, data privacy, auditability, governance frameworks

PROFESSIONAL EXPERIENCE

Lumenalta (formerly Clevertech)

Senior GenAI & ML Engineer

March 2024 – Present

Toronto, ON

- Architected and deployed **production multi-agent GenAI systems** using LangGraph, LangChain, and AWS Bedrock Agents — orchestrating multiple LLMs with specialized tools, memory, and retrieval across enterprise client environments in finance, energy, and e-commerce.
- Built **enterprise RAG pipelines** using AWS Bedrock Knowledge Bases and pgVector — ingesting and indexing large document corpora for semantic retrieval; integrated with foundation models for accurate, context-grounded LLM responses at scale.
- Designed **prompt engineering frameworks and reusable AI component libraries** — standardizing prompt design, evaluation, and versioning patterns across client teams to ensure consistent, production-reliable LLM outputs.
- Developed **FastAPI-based LLM service endpoints**, containerized with Docker, and integrated into orchestration systems — enabling scalable, maintainable AI delivery with CI/CD automation via GitHub Actions.
- Applied **AWS Bedrock Guardrails** in production for content filtering, PII redaction, and topic safety controls — ensuring responsible AI deployment meeting enterprise compliance requirements.
- Built **MLOps-grade pipelines** using MLflow, model registries, and Great Expectations for anomaly detection across 20,000+ tables on Apache Spark and Airflow with automated drift alerting.

LearnAIDaily

Machine Learning Engineer

October 2023 – March 2024

Toronto, ON

- Designed and built **end-to-end GenAI application pipelines** using LangChain and LlamaIndex — ingesting, transforming, and indexing diverse learning content for LLM retrieval and personalized AI delivery.

- Defined **MLOps standards for LLM inference**: reproducible training workflows, model versioning, inference monitoring, and deployment pipelines ensuring production reliability and SLA compliance.

Scotiabank

Data & ML Engineer, Data Platform

May 2022 – March 2024
Toronto, ON

- Led design of an **enterprise ML anomaly detection platform** — time-series models with scikit-learn and PySpark detecting data quality failures across **800+ production datasets**; reduced mean time to detect (MTTD) for critical incidents.
- Established **CI/CD and DevSecOps practices** for ML pipeline code promotion using GitLab CI, achieving **200% increase in deployment velocity**; built monitoring dashboards driving 80% improvement in stakeholder engagement.
- Authored **reusable ML testing frameworks** (Great Expectations, dbt tests, PySpark unit tests) and mentored engineers on ML software engineering best practices — modelling standards, test coverage, and deployment patterns.

Scotiabank

Software Engineer, Enterprise Data Lake

June 2021 – May 2022
Toronto, ON

- Modernised ETL pipelines from Hadoop/Hive to **Apache Spark and Kafka** for 20+ business clients; built reusable PySpark transformation libraries and SQL validation frameworks with schema enforcement and unit test coverage.

University of Toronto

Machine Learning Researcher — Cyber-Physical Systems Security Lab

Sep 2018 – Dec 2020
Toronto, ON

- Developed **LSTM-based deep learning models** for fault detection in smart grid datasets; applied cross-validation, ablation studies, and rigorous evaluation — achieving **1.3× performance improvement** over state-of-the-art baselines. Published MASc thesis and **4 peer-reviewed papers**.

SELECTED PROJECTS

- **Multi-Agent RAG Platform** (*Lumenalta*) — Designed and deployed a production multi-agent system coordinating retrieval, synthesis, and validation agents over enterprise document corpora using LangGraph and AWS Bedrock Agents. Integrated Bedrock Knowledge Bases for semantic search, Guardrails for content safety, and FastAPI with Docker for scalable LLM service delivery. Shipped to a live enterprise client with full GitHub Actions CI/CD.
- **Fraud Detection ML Pipeline** (*Scotiabank AI Hackathon*) — End-to-end Python pipeline: SQL feature engineering on transaction data, XGBoost ensemble classifier, SMOTE resampling for class imbalance, and precision-recall threshold calibration. Findings presented to senior leadership.

CERTIFICATIONS

Databricks Certified ML Associate — [Completed](#)

Databricks Certified Data Engineer Professional — [Completed](#)

Deep Learning Specialization, Coursera / [deeplearning.ai](#) — [Completed](#)

Professional Data Engineer, Google Cloud — [In Progress](#)

EDUCATION

Master of Applied Science (MASc), Electrical & Computer Engineering **2021**
University of Toronto • Applied ML for Cyber-Physical Security • 4 peer-reviewed publications

Bachelor of Technology (BTech), Electrical Engineering **2016**
IIT Roorkee — **Gold Medalist** • Summer Research: Carnegie Mellon University