Sanika Killekar

Boston, MA | (857) 294-8020 | killekar.s@northeastern.edu | GitHub | LinkedIn

SUMMARY

I build end to end data engineering and AI systems: I design reliable data pipelines in Python and SQL with Airflow, Spark, Kafka, and cloud storage, and I also build agentic AI and RAG applications that help teams search, summarize, and act on information with grounded outputs. I focus on production readiness through testing, CI/CD, monitoring, and secure deployments.

WORK EXPERIENCE

InfinityPool Finnotech Private Limited - Machine Learning Research Intern *May 2024 – May 2025*

- Led an end-to-end **agentic RAG** pipeline in **LangGraph** (routing, tool-calls, retries) to generate grounded next-day market summaries; processed **48k+ docs/day** and improved signal accuracy by **15%**.
- Served the LLM using **vLLM** on GPU and reduced GPU utilization by **28%** by eliminating redundant requests.
- Orchestrated ingestion with **Apache Airflow and Docker**; streamed feeds via **Kafka**; stored raw docs in **S3** and structured data in **PostgreSQL**, achieved **99.2%** successful DAG runs
- Built an internal analyst copilot using RAG with **ChromaDB** and tool-calling (SQL and document search) plus guardrails;
- supported **50+ daily users** and reduced repeat queries by **27%.**
- Deployed on **Kubernetes** and monitored with **Prometheus and Grafana**; tracked ingestion lag, RAG latency, and uptime.

Tata Institute of Fundamental Research (TIFR) - Computer Vision Research Intern *May 2023 – Apr 2024*

- Deployed a citation-first research copilot (**LangChain , FAISS, and GPT-4-turbo**) that searches, verifies, and summarizes evidence across **90K+ internal papers**, reducing literature review time by **35%**.
- Fine-tuned using **LoRA/QLoRA** to improve response consistency and reduce unsafe outputs; validated improvements with repeatable evaluation runs.
- Replaced **Azure Form Recognizer** with Unstructured.io for PDF parsing at scale, **lowering per-document processing cost**.
- Designed automated **ETL** pipelines (**Azure Data Factory and PySpark**) to ingest and clean heterogeneous sources (PDFs, patents, SQL DBs), reducing preprocessing time by **31%**.
- Containerized and deployed the RAG app on **AKS** using **Docker**, integrating authentication via **Azure Key Vault**.
- Automated **CI/CD** with **GitHub Actions** (unit tests), cutting deployment time from **45 to 10 mins**.

EDUCATION

Northeastern University, Boston, MA *Aug 2025 – May 2027*

Master of Science in Data Science

Relevant Coursework: Machine Learning, MLOps, Algorithms, Database Management Systems, Data Engineering

University of Mumbai, India *Aug 2021 – May 2025*

Bachelor of Engineering in Information Technology

TECHNICAL SKILLS

Programming: Python, SQL, Bash, R | **Operating Systems:** Linux, Windows | **Analytics:** Power BI, Tableau, Looker, Excel
Machine Learning: PyTorch, TensorFlow, Scikit-learn, XGBoost, Pandas, NumPy, NLP
LLMs & RAG: LangChain, LangGraph, OpenAI API, FAISS, ChromaDB, vLLM, FastAPI
Data Engineering: PySpark, Databricks, PostgreSQL, MySQL, ClickHouse, S3, Snowflake
Cloud & MLOps: AWS, Azure, GCP, Docker, Kubernetes, Apache Airflow, Kafka, GitHub Actions, Terraform, Prometheus, Grafana

PROJECTS

HubScout - Large-Scale Semantic Retrieval System

Python, FastAPI, LangChain, ChromaDB/FAISS, vLLM, GCP

- Built a semantic retrieval and RAG service over **350k+ records**; improved search relevance using embedding and reranking.
- Deployed the **FastAPI** service with GPU-backed inference (**vLLM**); added **agent behavioral regression tests** (tool-use accuracy, grounding rate, harmful-output checks) with red-team prompts to keep releases stable.

IMDb ML Data Pipeline

Azure Data Factory, Databricks (PySpark), Snowflake, dbt, Great Expectations

- Built Azure Medallion (Bronze/Silver/Gold) pipeline to load **180M+ IMDb rows** into Snowflake, cutting prep time by **65%**.
- Added dbt and Great Expectations checks and shipped **Power BI** dashboards, improving data reliability by **85%**.

PUBLICATIONS & INTELLECTUAL PROPERTY

Predictive Analysis Based on Leading Parameters for Stock Market, IEEE ICNTE-2026 Nov 2025

Crowd Flow Analyzer (Copyright Registered), Govt. of India Jul 2024