ANIKET CHAUHAN

AI Engineer | Applied GenAI & Backend Systems

+91 : 8169882434 aniketchauhan0608@gmail.com www.linkedin.com/in/aniket-chauhan-110b8727 Navi Mumbai, India.

SUMMARY

AI Engineer with **4+ years of industry experience** across software engineering and AI-driven systems, currently pursuing **M.Tech in Artificial Intelligence & Machine Learning from BITS Pilani**. Hands-on experience in **Generative AI, LLM fine-tuning, Retrieval-Augmented Generation (RAG), Vision-Language systems, computer vision, and end-to-end AI pipelines**. Strong background in integrating AI models into **enterprise and industrial applications**, deploying **open-source models**, and building scalable, production-ready systems with real-world business impact.

SKILLS

Applied AI / GenAI: RAG, VLMs, LLM Integration, Prompt Engineering, LoRA (applied), YOLO, OpenCV, ArcFace
Backend: Java, Spring Boot, REST APIs, Kafka, Elasticsearch, Microservices
AI/Data: Python, PyTorch, TensorFlow, Hugging Face, Pandas, NumPy
MLOps: Docker, CI/CD, Prometheus, Grafana, Git
Cloud/DB: AWS, GCP, PostgreSQL, MongoDB

KEY AI/ GEN AI PROJECTS

WalkSense – Vision-Language AI System

10/2025 - Present

WalkSense − Safety-First AI Navigation System for Visually Impaired Users

- Built a real-time **multimodal AI system** combining **YOLO-based** object detection, **Vision-Language Models (VLMs)**, and **LLM- based** reasoning for contextual scene understanding.
- Designed a **safety-isolated**, rule-based **perception layer** to ensure critical alerts remain **deterministic** and independent of generative models.
- Implemented **Vision-Language** and **LLM inference pipelines** with frame sampling and orchestration to control model usage and maintain low-latency performance.
- Integrated **Speech-to-Text (STT)** for voice commands and **Text-to-Speech (TTS)** for real-time audio feedback, enabling hands- free interaction.
- Performed **domain-specific fine-tuning** and **controlled retraining**, applying **prompt engineering** and response validation for grounded, reliable outputs.
- Established **CI/CD pipelines** for model and service deployment using **Git** and **Docker**.
- Implemented **performance benchmarking** and monitoring using **Prometheus** (latency, throughput, error rates) and **Grafana** dashboards.
- Optimized the system for **local** / **offline inference**, improving robustness and resource efficiency.
- **Tech:** Python, PyTorch, Hugging Face, YOLO, Vision-Language Models, LLMs, OpenCV, FastAPI, Docker, Prometheus, Grafana, STT/TTS

Facial Recognition System (Edge AI)

05/2025 - 08/2025

Facial Recognition System – Workforce & Weighment Verification

- Built a **facial recognition–based** identity verification system for **daily wage worker attendance** and weighment validation.
- Used **ArcFace embeddings** and **cosine similarity** to ensure the same individual performed assigned weighment tasks, preventing proxy attendance.
- Integrated **AI verification with weighment** and **transaction records** for auditing and reporting.
- Designed the **system to handle real-world conditions** and scale to over **10,000 identities** with low latency.
- Deployed models using **on-device** inference (**TensorFlow Lite**) to enable offline, real-time verification.
- **Tech:** Python, TensorFlow Lite, ArcFace, OpenCV, Android, REST APIs

Poultry Disease Severity Detection (AI Mobile App)

11/2024 - 02/2025

AI-Assisted Poultry Disease Severity Detection (Mobile Application)

- Developed an **AI-enabled mobile application** for **poultry disease severity detection** using image-based analysis.
- Implemented **computer vision** and **machine learning-based classification** to assess visible disease indicators and severity levels in poultry.
- Designed **preprocessing pipelines** to handle **real-world image** variations such as lighting, posture, and camera quality.
- Integrated **backend services** to generate **severity scores** and **provide treatment guidance** for operational decision support.
- Enabled mobile-first inference workflows, ensuring quick feedback for field usage.
- **Tech:** Python, computer vision, machine learning classification (SVM), OpenCV, Android, REST APIs.

EXPERIENCE

Senior Software Engineer
02/2024 - 01/2025

Long Health

US-Based company focused on healthcare technology

- Designed and implemented an LLM-powered **Retrieval-Augmented Generation (RAG)** system to enable contextual retrieval of healthcare guidelines, referral workflows, and operational knowledge.
- Built document ingestion, **chunking**, and **vector-based semantic search** pipelines for grounded, **non-hallucinated responses**.
- Integrated GenAI services with **FastAPI** and **Spring Boot**, exposing AI insights to enterprise applications.
- Implemented real-time data pipelines using **Kafka** and **Elasticsearch**, reducing referral processing latency by approximately 40%.
- Collaborated with **cross-functional teams** to ensure privacy-aware, **reliable GenAI** adoption in production systems.
- **Tech:** Java, Spring Boot, Python, FastAPI, Kafka, Elasticsearch; RAG, LLM Integration, Docker, CI/CD

SDE II
07/2023 - 01/2024

Esmito Solutions Pvt. Ltd

A tech company specializing in EV Battery Swapping G Charging Infrastructure

- Revamped a traditional, legacy **battery-swapping** system into a **modern**, scalable backend platform, improving reliability and performance across multiple charging locations.
- Contributed to **AI-enabled battery analytics** systems for **EV battery-swapping stations** deployed across multiple sites.
- Implemented **battery health** and **eligibility logic** using **telemetry data** (charge cycles, voltage, temperature) to ensure safe and compliant battery release.
- Built **real-time data ingestion**, **preprocessing**, and **aggregation pipelines** for **battery** and **charging-station analytics**.
- Integrated **backend services** to validate **user authorization**, **battery status**, and station readiness before swap execution.
- Developed **analytics** and **control APIs** consumed by **mobile** and **web applications** used by large partner fleets.
- Designed and integrated an **internal GenAI-based RAG** system to **query battery SOPs**, station manuals, and incident-resolution workflows for operational decision support.
- Implemented **document ingestion, chunking, and vector-based semantic search for reliable, grounded responses**.
- Improved system stability through **CI/CD pipelines**, **automated testing**, and performance optimization.
- **Tech:** Java, Spring Boot, Python, FastAPI, SQL, Docker, CI/CD, AWS, RAG, Vector Search, Telemetry Analytics.

Software Development Engineer (R & D)
06/2021 - 07/2023

Ignisnova Robotics Pvt. Ltd.

Innovation in robotics and AI technologies (AgriTech)

- Designed and developed an **AI-driven analytics** platform for **crop monitoring** and **disease surveillance** using sensor data, weather APIs, and image inputs.
- Built **data ingestion**, **validation**, and **preprocessing pipelines** combining soil parameters, weather conditions, and crop images.
- Implemented **computer-vision–based disease detection** and severity classification using **ML** and **image feature extraction**.
- Developed risk scoring logic to identify early crop stress and potential disease outbreaks.
- Generated ML-assisted **recommendations** for **irrigation**, **fertilization**, and **pesticide usage**, **improving planning efficiency by approximately 80%**.
- Exposed **analytics and AI insights through REST APIs** consumed by **mobile** and **web dashboards**.
- Deployed scalable backend services on cloud infrastructure, ensuring reliability and high availability.
- **Tech:** Python, Java, Spring Boot, Android, REST APIs, SQL, GCP; Computer Vision, ML Classification, Pandas, NumPy

EDUCATION

MTech in Artificial Intelligence & Machine Learning
Pilani

BITS Pilani

05/2024 - Present

BE in Computer Engineering
Navi Mumbai

SIGCE

01/2017 - 06/2021

RESEARCH & ACHIEVEMENTS

Research Publication

Intelligent Traffic Lights using Li-Fi Technology — published in the **International Journal of Emerging Technologies and Innovative Research (IJETIR)**.).

Academic Citation

The research has been **cited in the book *World Smart Cities*** by a Portuguese professor and is available in libraries of leading universities, including **Stanford University and Cornell University**.

Hackathon Winner

1st place – Gupshup Hackathon (out of 28 teams)