Zayed Ansari

Mumbai, Maharashtra
+91-9356635531 | zayed.ansari04@gmail.com | LinkedIn | GitHub | Twitter(X)

Education

D Y Patil College of Engineering — **Pune, Maharashtra**
B.Tech Artificial Intelligence and Data Science | ***CGPA: 9.15*** — *November 2022 – July 2026*

Poddar Brio International School — **Mumbai, Maharashtra**
XII — *March 2020 – August 2021*

Technical Skills

Languages: Python, C++, HTML/CSS, JavaScript
Generative AI: LangChain, RAG Systems, Vector Databases, Fine-Tuning
Frameworks: PyTorch, TensorFlow, Scikit-Learn, Hugging Face Transformers
Development and DevOps: Docker, Flask, Fast API, Git, CI/CD, Hugging Face Spaces
Database & Data: PostgreSQL, PySpark, Pandas
Cloud: Amazon Web Services

Experience

Python Developer Intern — **Jan 2026 – Present**
Helo.ai by Vivaconnect — *Mumbai, Maharashtra*

- Developed a PySpark-based behavioral profiling pipeline ("WHEN" model) with Dockerized deployment and clear technical documentation to support marketing analytics..
- Built a mini Customer Data Platform (CDP) for ML-driven customer segmentation using RFM features and K-means clustering on 500K+ customer records.
- Engineered data pipelines in Python and PostgreSQL to compute customer-level metrics (recency, frequency, monetary, churn risk) for targeted campaigns.
- Trained and evaluated churn and value scoring models (e.g., Random Forest) to prioritize high-risk, high-value segments for retention.

Projects

Multi-Document RAG with Source Citations | *LangChain, Streamlit, Gemini API* [**Source Code**] — **January 2026**

- Built production-ready Retrieval-Augmented Generation system enabling multi-document querying across PDFs with automatic source citations, processing 27-page documents with sub-second response times.
- Implemented vector-based semantic search using ChromaDB embeddings and LangChain orchestration, achieving 90%+ answer relevance through context-aware retrieval.
- Integrated source attribution mechanism tracking document origins and page numbers, enhancing transparency and enabling users to verify AI-generated responses with original sources.
- Evaluated local LLM (Ollama) vs cloud API trade-offs, optimizing for deployment feasibility and response quality by switching to Gemini API, improving answer coherence by 40%.

Sonnet Generator | *GPT-2, NLP, Fine Tuning* [**Source Code**] [**Video Demo**] — **May 2024**

- Engineered creative AI system using fine-tuned GPT-2 to generate Shakespearean sonnets, earning 90% user satisfaction via automated rhyme and meter preservation.
- Optimized loss functions for poetic structure, surpassing baseline model by 15% in coherence metrics (BLEU/perplexity) while generating 1,000+ outputs.
- Deployed interactive web UI with Flask and Docker, slashing content creation time by 70% for educators and enabling seamless access for non-technical users.
- Collaborated with beta testers to iterate on feedback, integrating RAG systems for enhanced contextual relevance.

Real-Time Pothole Detection System | *Ensemble Learning, Transfer Learning, Edge AI* — **Feb 2024**

- Developed ensemble CNN model (Xception + InceptionV3) achieving 90-95% accuracy in classifying road conditions, laying groundwork for automated road maintenance systems that could reduce manual inspection costs.
- Curated custom dataset of 3,500+ images from field captures, video extraction, and online sources, addressing data scarcity challenges in road infrastructure monitoring applications.
- Reduced model training iterations by 60% through strategic transfer learning approach, progressing from under-performing baseline CNN to optimized ensemble architecture.
- Deployed model to Raspberry Pi with real-time camera inference, demonstrating feasibility of low-cost ($35) edge AI solution for municipal road quality monitoring.