

# FARHATHUL AFRAH S

## AI/ML Engineer, Data Scientist

Chennai, India | [afrahthahir@gmail.com](mailto:afrahthahir@gmail.com) | +919840943720 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

### Professional Summary

AI/ML Engineer with 3 years of experience building, deploying, and optimizing LLM-powered, RAG-based, and cloud-native ML systems. Proficient in Python, PyTorch, TensorFlow, LangChain, Pinecone, MLFlow, and AWS, with expertise in fine-tuning, MLOps automation, and end-to-end AI deployment. Proven track record of transforming research prototypes into production-grade solutions that deliver measurable business impact.

### Skills

---

- **Languages:** Python, C++, SQL, React js, Node js, Javascript
- **Libraries & Tools:** Pandas, Numpy, Matplotlib, Seaborn, Plotly, Scipy, Sckitlearn, Spacy, NLTK, HuggingFace, LangChain, TensorFlow, Keras, PyTorch, Scikit-learn, FAISS, Pinecone, ChromaDB, LoRA Fine-tuning, NLTK, SpaCy, Sentence Transformers, Flask, Beautiful soup, n8n, Matplotlib, Seaborn, Zoho Analytics, Streamlit
- **MLOps & Cloud:** MLFlow, AWS (Lambda, EC2, S3, ECS, EFS), FastAPI, Docker, GCP
- **Data Wrangling:** Data generation, Data Extraction, Data Cleaning, Exploratory Data Analysis, Feature Engineering
- **ML Skills:** Data Modeling, Clustering & Classification, Quantitative Analysis, Regression, Model Validation, Model deployment, CNN & RNN, LSTM, BERT, Transformers.

### Work Experience

---

#### Agentz, Chennai | *Software Developer* | 01/2023 – Present

- **LLM & RAG Development:** Integrated OpenAI LLMs with Pinecone and AWS Lambda, reducing response latency by 20%.
- **Voice AI Deployment:** Implemented Retell voice agents for customer support, cutting average call time by 25%.
- **Agentic AI Automation:** Built autonomous multi-step workflows using n8n for tool-based task execution.
- **Analytics Automation:** Created Zoho Analytics pipelines, accelerating reporting by 40% and reducing manual effort by 30%.
- **Full-Stack Integration:** Connected backend ML APIs with frontend features, boosting user engagement by 15%.

**Key Technologies:** Python, AWS, LangChain, Pinecone, FastAPI, Docker, MLFlow

## Education

---

University College of Engineering, Kancheepuram  
**B.E. Computer Science Engineering**

2018-2022  
**CGPA:8.56/10**

## Languages

---

**English** – Professional Proficiency

**Arabic** - Basic

**Hindi** – Professional Proficiency

**Tamil** – Native

## Projects

---

### **Cold Email Generator (LLM + RAG)** — *AI Outreach Tool*

- Built a personalized email generator using LangChain, ChromaDB, and LLaMA-3.1 with 90%+ relevance. Deployed a retrieval-augmented pipeline for context-aware content generation.

### **News Article Chatbot (RAG)** — *Contextual Q&A over Live News*

- Created a real-time news chatbot using Google Palm LLM, LangChain, and FAISS for dynamic retrieval and QA.

### **Log Classification (Hybrid ML + LLM)** — *Anomaly Detection Pipeline*

- Combined Regex, DBSCAN, Sentence Transformers, and Logistic Regression to detect system anomalies proactively, deployed ML pipeline on AWS for automated monitoring.

### **LLM Fine-Tuning with LoRA** — *Text Classification on arXiv Dataset*

- Fine-tuned LLaMA models via LoRA for efficient domain adaptation and improved classification accuracy, deployed cloud-based inference endpoints for real-time usage.

### **Potato Disease Prediction (CNN + Cloud)**

- Deployed a TensorFlow CNN on AWS EC2 and GCP to detect leaf diseases (~85% accuracy) for real-time field use.

### **Traffic Violation Detection (YOLOv5 + ANPR)**

- Built a real-time helmet and triple-riding detector using YOLOv5 and PaddleOCR; achieved 85% mAP on custom dataset.

### **Twitter Sentiment Analysis API (Transformer + FastAPI + Docker)**

- Fine-tuned TFDistilBertForSequenceClassification on the Twitter Entity Sentiment Analysis Dataset for 2-class sentiment analysis, achieving 93.3% validation accuracy.
- Deployed as a Dockerized FastAPI service with Uvicorn + Gunicorn asynchronous workers, handling 500+ concurrent requests, and stress-tested with Locust for high-throughput, low-latency performance.