# DHAYANIDHI PALANI

**Phone :** +91 8148024396 | **Email :** dhayanidhi962004@gmail.com | **Location :** Dharmapuri
**LinkedIn** : Dhayanidhi P | **GitHub** : Dhayanidhi-96

## PROFILE SUMMARY

AI/ML Engineer and M.Sc. Data Science candidate with hands-on experience building and deploying end-to-end ML and LLM systems across multiple internships. Currently fine-tuning Gemma 4 for Tamil regional language understanding at ViteTech. Shipped two live AI applications including a production RAG chatbot built with LangChain, LLaMA 3.1, FAISS, and Groq API. Skilled in Python, deep learning, RAG pipelines, LLM fine-tuning (LoRA/QLoRA), REST API development, and cloud deployment. Seeking AI/ML Engineer roles at startups and AI-first companies.

## SKILLS

| | |
|---|---|
| Language | Python (Data Science & ML), SQL |
| AI / LLMs | LangChain, LLaMA 3.1, Groq API, Prompt Engineering, RAG, HuggingFace Transformers & Embeddings, FAISS, History-Aware Retrieval,LoRA,QLoRA Fine Tuning |
| Deep Learning | CNN, RNN, LSTM, Transfer Learning — TensorFlow, Keras, PyTorch |
| Machine Learning | Regression, Classification, Clustering, Feature Learning Engineering, Model Evaluation, MLflow |
| Data Science | EDA, Data Wrangling, Preprocessing, Visualization — Pandas, NumPy, Matplotlib, Seaborn |
| Deployment | FastAPI, Streamlit, Docker, Render, Hugging Face Spaces |
| Databases | MySQL, Microsoft SQL Server, SQLite , Redis |
| Tools | Git, GitHub, Jupyter, Google Colab, VS Code |

## Experience

**AI Intern | Bipolar Factory |** **July 2026 – Present**

- Built an end-to-end CCTV image processing pipeline for retail stores, using remote device access to ingest live camera feeds and applying ROI-based cropping to isolate specific store zones for analysis
- Designed a Kafka-based streaming architecture to route full-frame and zone-cropped images from edge devices to a VLM inference service, enabling automated visual judgment of store conditions
- Engineered the inference-to-action loop: parsed structured JSON outputs from the VLM, pushed results to a monitoring dashboard, and forwarded flagged events to a downstream production Kafka pipeline for automated store-ops alerts (e.g., shelf/product misplacement)
- Fine-tuned a domain-specific vision model using LoRA/QLoRA on synthetic CCTV-style imagery generated with an open-source Flux-based diffusion pipeline, improving model robustness for retail surveillance use cases
- Used Claude Code to accelerate pipeline development across Python-based Kafka producers/consumers, VLM integration, and model fine-tuning workflows

**AI/ML Intern — Vite Tech** **April 2026 – June 2026**

- Fine-tuning Gemma 4 for Tamil regional language support, conducting end-to-end research on LoRA/QLoRA adaptation strategies for low-resource Indic language domains.
- Performing large-scale data preprocessing and quality checks on Tamil text corpora, including deduplication, normalization, and domain filtering for instruction-tuning datasets.
- Researching and documenting fine-tuning pipelines using HuggingFace PEFT and TRL libraries; using Claude Code for accelerated implementation and research workflows.
- Driving daily research cycles covering dataset curation, tokenizer analysis, training configuration, and evaluation metrics for multilingual LLM fine-tuning.

**Data Science Intern — PK Software Solutions** **Dec 2025 – April 2026**

- Built an AI-powered research paper discovery system with semantic search using FAISS, Sentence-Transformers (MiniLM-L6), and Mistral 7B, with a FastAPI backend and PostgreSQL database.
- Implemented Redis caching layer achieving sub-500ms query performance and designed RESTful search endpoints with Pydantic schema validation across 13,000+ research papers.
- Gained full-stack development experience covering Python backend, PostgreSQL schema design, vector indexing, and API development.

## PROJECTS

**AI STUDY ASSISTANT — RAG-Powered PDF Chatbot** **2026**

Python | LangChain | LLaMA 3.1 | Groq API | FAISS | Streamlit | Docker
Live Demo | Github

- Architected and deployed a production RAG chatbot on Hugging Face Spaces, enabling students to query PDFs conversationally using semantic retrieval over FAISS vector stores with HuggingFace sentence embeddings.
- Implemented a history-aware retriever using LangChain's create_history_aware_retriever to handle multi-turn conversations and maintain context across follow-up questions.
- Optimized inference latency using Groq API's LPU-accelerated LLaMA 3.1, achieving near-instant response times compared to standard CPU/GPU inference endpoints.
- Containerized the application with Docker for reproducible deployment; launch post generated 6,800+ LinkedIn impressions and reached 3,900+ members organically within 24 hours of release.

**AI-POWERED RESEARCH PAPER DISCOVERY & SEMANTIC SEARCH SYSTEM** **2026**

Python | FastAPI | PostgreSQL | FAISS | Sentence-Transformers | Mistral 7B | Redis | SQLAlchemy

- Built an end-to-end research paper discovery system with an arXiv scraper, PostgreSQL relational schema, and 384-dimensional vector embeddings using Sentence-Transformers MiniLM-L6 stored in a FAISS index for semantic similarity search.
- Developed a FastAPI backend with Pydantic schemas and search endpoints, integrating an Ollama/Mistral 7B pipeline to perform batch analysis and extract structured insights across 13000 research papers.
- Implemented Redis caching layer achieving sub-500ms response times for repeated semantic search queries.

---

## CERTIFICATIONS

- Machine Learning Specialization — DeepLearning.AI / Coursera (Andrew Ng)
- Python for Data Science, AI & Development — IBM / Coursera
- Google Data Analytics Certificate — Google / Coursera

---

## EDUCATION

| | | | |
|---|---|---|---|
| Periyar University | M.Sc. Data Science | CGPA: 8.21 | 2024 – 2026 |
| Presidency College, Chennai | B.Sc. Statistics | CGPA: 8.01 | 2021 – 2024 |
| Senthil Matric Higher Secondary School | HSC | 90% | 2019 – 2021 |
| Senthil Matric Higher Secondary School | SSLC | 88% | 2018 – 2019 |

---

## ACHIEVEMENTS

- Deployed two live production AI/ML applications accessible to real users, demonstrating full end-to-end ownership from data collection to cloud deployment.
- AI Study Assistant launch organically reached 10,000+ LinkedIn impressions and 6,900+ members within 24 hours with zero paid promotion.
- Currently contributing to multilingual LLM fine-tuning for Tamil, an under-resourced language with 80M+ speakers, at ViteTech