

Harish Kumar S

AI/ML Engineer

harishkumar56278@gmail.com | +91 73973 27957

Ranipet, Tamil Nadu, India | Chennai, India

<https://www.linkedin.com/in/harish-kumar-s-b0a0472b1/>

<https://github.com/Harish-nika>

<https://harish-nika.github.io/>

PROFESSIONAL SUMMARY

Integrated MTech in Data Science (VIT, 2025) AI/ML engineer building production AI systems at FactEntry Data Solutions (A SIX Company). Experienced in MLOps, Kubernetes, GitOps, RAG pipelines, computer vision, and scalable backend engineering. Delivered enterprise-grade AI platforms using FastAPI microservices, Redis-based distributed processing, and cloud-native deployment architectures.

- Built and deployed a production bond-extraction platform on K3s, leveraging 80+ CPU cores with Redis Streams and human-in-the-loop validation workflows.
- Implemented end-to-end GitOps delivery pipelines using GitLab CI, Helm, and Argo CD for automated staging deployments.
- Developed RAG-powered document intelligence systems using FAISS, Groq LLaMA, Streamlit, and FastAPI.
- Published the language-processing-tool package on PyPI for reusable NLP workflows.
- Trained and deployed YOLOv11 computer vision models on Roboflow for dental image analysis applications.

CORE COMPETENCIES

AI/ML Engineer | Machine Learning | Deep Learning | Computer Vision | NLP | RAG | Python | FastAPI | Kubernetes | K3s | Docker | Redis | PostgreSQL | GitLab CI | Argo CD | Helm | GitOps | MLOps | Roboflow | YOLO | FAISS | LangChain | Streamlit

EXPERIENCE

Software Developer Trainee (Probation)

FactEntry Data Solutions - A SIX Company | Dec 2025 - Present

- Shipped Data Mining Engine (DME) modules on K3s stage: corporate/municipal bond extraction, validation UI, and cluster operations portal.
- Implemented Redis-stream job dispatch with multi-node worker throughput and GitOps delivery (GitLab CI image build -> Helm chart promote -> Argo CD).
- Built FastAPI microservices, live WebSocket job telemetry, role-based CPU tiering, and domain-restricted enterprise auth.

MLOps and Data Science Intern

FactEntry Data Solutions - A SIX Company | Nov 2024 - May 2025

- Published language-processing-tool on PyPI (PDF language detection, OCR, document clustering) as self-driven internship contribution.
- Built Fixed Income Expert RAG system (FAISS, Groq LLaMA 3, Streamlit/FastAPI) as personal contribution during internship.

Associate Software Engineer Trainee Intern

MIMASOFT Technologies | Jun 2025 - Oct 2025

- Developed Dental ML Project: Roboflow YOLOv11 tooth and panoramic X-ray models with inference CLI.
- Contributed to backend integration, API workflows, and production-oriented software development practices.

EDUCATION

Integrated MTech - CSE (Data Science), Vellore Institute of Technology (2020 - 2025) - CGPA 7.8

Class XII, THIRU G V C Higher Secondary School (2020) - 79%

Class X, St Joseph Matriculation Higher Secondary School (2018) - 87.8%

TECHNICAL SKILLS

AI/ML: Machine Learning, Deep Learning, Generative AI, Computer Vision, NLP, RAG, FAISS, Sentence Transformers, Ollama, Roboflow, YOLOv11, Supervision

Engineering: Python, FastAPI, Streamlit, SQL, Docker, Kubernetes (K3s), Redis, PostgreSQL, LangChain, PyMuPDF, OpenCV

DevOps/GitOps: GitHub, GitLab CI, Argo CD, Helm, systemd, GitHub Actions, GitOps

PROJECTS

Data Mining Engine (DME)

Where: FactEntry Data Solutions (A SIX Company) - Software Developer Trainee, probation

Enterprise bond extraction on K3s - UI, API, Redis queue workers, GitOps CI/CD, validation workspace.

Tech: K3s, FastAPI, Redis, PostgreSQL, Argo CD

<https://harish-nika.github.io/dme-project.html>

Recruiting Agent - AI-Powered ATS

Where: Independent R&D - Agents Monorepo

Multi-step LLM ATS - resume OCR, ChromaDB JD matching, verification, suspicion scoring, and HR feedback loop.

Tech: FastAPI, React, SQLite, ChromaDB, Ollama, Groq

<https://harish-nika.github.io/recruiting-agent.html>

<https://github.com/Harish-nika/Agents>

Fixed Income Expert System

Where: FactEntry internship - personal contribution

RAG chatbot for financial bond PDFs - upload, embed, retrieve, LLaMA 3 answers via Groq.

Tech: FastAPI, Streamlit, FAISS, Groq, SBERT

<https://harish-nika.github.io/fexpert.html>

<https://github.com/Harish-nika/FExpert>

Cybersecurity Content Moderator

Where: VIT - Capstone project (academic)

Multimodal hate-speech and harmful content classification (text, image, PDF).

Tech: LLM, Streamlit, Docker, FastAPI

<https://harish-nika.github.io/content-moderator.html>

<https://github.com/Harish-nika/Content-moderator-image>

Language Distribution & Document Clustering Tool

Where: FactEntry internship - self contribution

PyPI-published PDF language detector, OCR, and clustering pipeline (v0.2.7).

Tech: Python, Tesseract, LangDetect, PyPI

<https://harish-nika.github.io/lang-tool.html>

https://github.com/Harish-nika/language_processing_tool

MedBot (Medical RAG Assistant)

Where: VIT - J Component, Soft Computing course

Medical PDF Q&A with SBERT + FAISS retrieval and Groq LLaMA 3 grounded responses.

Tech: Streamlit, FAISS, Sentence Transformers, Groq

<https://harish-nika.github.io/medbot.html>

<https://github.com/Harish-nika/medbot>

Dental ML Project

Where: MIMASOFT Technologies - trainee internship

YOLOv11 tooth and panoramic X-ray detection - Roboflow fine-tuning and inference CLI.

Tech: Roboflow, YOLOv11, inference-sdk, Supervision

<https://harish-nika.github.io/dental-project.html>

https://github.com/Harish-nika/Dental_ML_project

Disease Prediction using Ensemble Learning

Where: SmartInternz / Smart Bridge - Applied Data Science externship

Ensemble ML model for clinical trial disease prediction with MongoDB Atlas pipeline.

Tech: Python, Random Forest, SVM, MongoDB

<mailto:harishkumar56278@gmail.com?subject=Disease%20Prediction%20Project>

Real-Time Room Temperature & Humidity Visualization (AR)

Where: VIT - J Component, ARVR course

MQTT telemetry with Unity AR overlay and ML-driven climate insights.

Tech: Unity, MQTT, AR, Python

<mailto:harishkumar56278@gmail.com?subject=AR%20Visualization%20Project>

Extractive Summarization & Information Retrieval

Where: JP Infotech

NLP web scraping with English-to-SQL query translation and 95% extraction accuracy.

Tech: Python, NLP, SQLite, MySQL

<mailto:harishkumar56278@gmail.com?subject=Summarization%20Project>

Application Tracking System

Where: JP Infotech

Resume analyser with OCR, keyword suggestions, and n-gram typing assist vs job descriptions.

Tech: Python, PyTesseract, OpenCV, NLP

<mailto:harishkumar56278@gmail.com?subject=Application%20Tracking%20System>

CERTIFICATIONS

- Microsoft Certified: Azure AI Fundamentals

https://www.credly.com/badges/da3ef4b3-03a0-4560-95b6-5941473704de/public_url

- Applied Data Science (SmartInternz + Google Developers)

https://drive.google.com/file/d/126d8VTJ6y-X-t6_S_KOD2gYttWQhqe2T/view

- Data Analysis and Visualization with Power BI

<https://www.coursera.org/account/accomplishments/verify/YKAC25V5T6WJ>

- Preparing Data for Analysis with Microsoft Excel

<https://www.coursera.org/account/accomplishments/verify/F3SR649G78ZN>

- Harnessing the Power of Data with Power BI

<https://www.coursera.org/account/accomplishments/verify/CTOOZYTXL926>

- Extract, Transform, and Load Data in Power BI

<https://www.coursera.org/account/accomplishments/verify/5VXS5AHV12DN>