

KF Surya

Phone: +91 8848617208 | Email: 22ad058@kpriet.ac.in | LinkedIn: [linkedin.com/in/kf-surya](https://www.linkedin.com/in/kf-surya) | GitHub: github.com/Surya-KF
Website: surya-kf.github.io/Portfolio/

Professional Summary

Enthusiastic B.Tech. student specializing in AI and generative AI with a passion for solving complex problems using machine learning, NLP, and computer vision technologies. Skilled in developing advanced AI applications, optimized RAG systems, and deploying AI-driven solutions in healthcare and other domains. I am seeking opportunities to leverage my skills in AI to drive innovation.

Work Experience

AI Intern

El Systems, Vaishali, Ghaziabad, Delhi NCR, India

Jul 2024 - Aug 2024

- Contributed to the development of advanced AI solutions, focusing on RAG systems and medical question-answering applications.
- Developed the Medi-Saga RAG Chatbot Application, utilizing RAPTOR indexing with Milvus vector database and Llama2.
- Engineered a High-Performance CPU RAG System with quantized Llama2 models, optimized for CPU deployment, increasing efficiency.

ML Intern

Yanne Technologies, Coimbatore, Tamil Nadu, India

Dec 2023 - Jan 2024

- Developed a LangChain-based Medical Assistant app using OpenAI GPT-3.5 for accurate medical information retrieval.
- Created a Checkbox Detection app using OpenCV for Vizhi EyeCare NGO, optimizing data processing tasks.

Projects

Medi-Saga - RAG Chatbot Application

- Built a medical chatbot using RAPTOR indexing, Milvus vector database, and Llama3 models, providing efficient medical query answers.
- Developed features like PDF text extraction, query expansion, and deployed a user-friendly interface with Streamlit.
- GitHub: <https://github.com/Surya-KF/MediSaga-RAG-Chatbot-Application>

High-Performance CPU RAG System with Quantized Llama2

- Engineered a RAG system optimized for CPU performance using quantized Llama2 models, achieving high efficiency without GPU.
- GitHub: <https://github.com/Surya-KF/High-Performance-CPU-RAG-System-with-Quantized-Llama2>

Automatos - Automated Surveillance and Messaging System

- Developed an automated surveillance system using Python and YOLO models for real-time human detection and alert notification.
- GitHub: <https://github.com/Surya-KF/AUTOMATOS-Automated-Surveillanceand-Messaging-System>

Churn Prediction in Telecom Sector

- Built a customer churn prediction app using Python, Scikit-learn, and Streamlit, transforming raw data into actionable insights.
- Streamlit: <https://chrunpredictionintelecomsector.streamlit.app/>

Technical Skills

- Programming Languages:** Python, JavaScript, Java, C
- Data Analysis:** NumPy, Pandas, R
- Visualization:** Matplotlib, Seaborn, Plotly
- AI & ML:** Supervised Learning, Unsupervised Learning, Neural Networks, CNNs
- NLP & RAG:** LangChain, GPT-3.5, Llama2, Sentence-Transformers, Milvus, Faiss-CPU.
- Tools:** TensorFlow, PyTorch, Milvus, Streamlit, Chainlit, MongoDB
- Computer Vision:** OpenCV, YOLO models, Image Processing

Certifications and Awards

- Best Paper Award (AMSAS-2022), Certified by **DRDO and Department of Chemistry-KPRIET**
- Machine Learning with Python, **IBM** (with **Honors**)
- Exploratory Data Analysis, Coursera
- Python for Data Analysis, Coursera
- Git and GitHub, 365 Data Science

Education

- B.Tech. in AI and Data Science** | KPR Institute of Engineering and Technology
 - 2022 - Present | CGPA: [7.87]
- Intermediate in Computer Science** | National Infotech College-Nepal
 - 2018 – 2020 | CGPA: [3.01 out of 4]