Lay Sheth

+91 7000035904 - laysheth1@gmail.com - LinkedIn:Cloaky233 - GitHub:Cloaky233 - Portfolio: cloaky.works

EDUCATION

VIT Bhopal University **09/2022 – 05/2026**

B.Tech in Computer Science - 8.84 GPA *Bhopal, India*

TECHNICAL SKILLS

- **Languages:** Python (Advanced), Rust, TypeScript/JavaScript
- **Frameworks:** PyTorch, TensorFlow, Keras, Burn, Hugging Face Transformers, GitHub Models, Google ADK, MCP SDK
- **Frameworks:** n8n,Streamlit, Tauri, FastAPI, Flask, Next.js

WORK EXPERIENCE

ThePreProdCorp **07/2024 – 12/2024**

Machine Learning Engineering Intern *Bengaluru, India (Remote)*

- Spearheaded AutoML solutions for automated model selection and hyperparameter tuning, reducing model development time by 40%
- Mapped and deployed end-to-end data pipelines integrating streaming data, real-time visualization, and model deployment
- Developed RAG systems with open-source LLM (Mistral), achieving 85% accuracy in context retrieval
- Collaborated with cross-functional teams to optimize data flow and enhance model serving architecture

PROJECTS

Retrieval Augmented Generation with Query Decomposition and GitHub Models **05/2025**

- Engineered production-ready RAG system using GitHub Models API, ChromaDB vector store, and HuggingFace embeddings with modular architecture supporting 40+ AI models including GPT-4, Llama, and Cohere
- Implemented intelligent query decomposition pipeline with streaming responses and async processing, automatically breaking complex queries into maximum 5 focused sub-queries for comprehensive analysis
- Developed enterprise-grade system with dependency injection patterns, real-time streaming capabilities, and comprehensive error handling using Python 3.13+ and UV package manager

High-Performance Web Scraping System: Python vs Rust Implementation **07/2024**

- Architected dual web scrapers using Python (BeautifulSoup4, requests) and Rust (Tokio, reqwest) with custom headers and rate limiting to extract product data from Amazon listings
- Optimized concurrent scraping using Rust's async/await patterns with Tokio runtime, implementing safe multi-threading and proper error handling for reliable data collection
- Delivered CSV data pipeline with custom serialization in both languages, demonstrating performance comparison between interpreted and compiled approaches for large-scale data processing

Distributed Real-time Image Classification System with Apache Kafka **08/2024**

- Collaborated with 1 team member to develop distributed image classification pipeline using Apache Kafka, TensorFlow, and Keras with pre-trained CNN model for cat/dog classification
- Established concurrent message processing architecture with Kafka producers, consumers, and topic partitioning using 3 main components: receiver-predictor, sender, and result consumer
- Implemented CNN architecture with 4 convolutional layers processing 256x256 grayscale images, integrating Python logging and Kafka metrics for real-time monitoring and error handling

CO-CURRICULAR

Smart India Hackathon(SIH) **12/2024**

National Level Hackathon Finalist *Team of 6*

- Pioneered cost-effective myoelectric prosthetic hand achieving 92% cost reduction (Rs 12,500) with companion mobile app processing EMG signals from dry electrodes, enabling real-time gesture recognition with 85% accuracy

PreProdCorp Buildathon **12/2024**

College Hackathon Winner *Team of 3*

- Pioneered an automated machine learning (AutoML) platform leveraging Streamlit and PyCaret, accelerating model training cycles by 75% and enabling data scientists to build models easily