Swaraj Khan P

+91 8618893815 swarajkhan2003@gmail.com LinkedIn Github LeetCode Portfolio YouTube

I build end-to-end AI systems that think, learn, and scale—engineering LLM-driven agents, RAG pipelines, and real-time data workflows that bridge research-grade innovation with production-level impact.

Education

Dayananda Sagar University, Bengaluru **2021 - 2025**
B.Tech in Computer Science & Engineering *7/10 CGPA*

The Amaatra Academy, Bengaluru **2020 - 2021**
Class 12 PCMC *79%*

Work Experience

Draconic AI **Oct '24 – Present**
Machine Learning Engineer — Founding Intern *In-Office Bangalore*

- Architected real-time trading data pipeline using FastAPI and Python, processing OHLCV data from Zerodha and Dhan platforms, achieving 99.9% uptime with dual-storage implementation using Redis and PostgreSQL (Supabase)
- Developed 15+ advanced trading computation modules in Python, including candlestick pattern recognition and options chain analytics, reducing algorithmic trading decision latency by 40%
- Designed secure authentication system integrating Google OAuth and WhatsApp OTP, supporting 1000+ concurrent users with Supabase backend infrastructure
- Engineered multi-agent AI systems using CrewAI, Agno, and Smol frameworks with Anthropic Claude and HuggingFace models, implementing comprehensive monitoring system that reduced token consumption by 25%
- Built dual-purpose RAG systems processing 60+ trading books and user-uploaded documents, achieving 97% query accuracy with LangChain and vector databases
- Optimized LLM performance through systematic prompt engineering with PromptLayer, improving response quality by 30% while reducing average response time to sub-150ms

Nokia **Mar '24 – July '24**
Chatbot Development — Intern *Hybrid Bangalore*

- Created AI-powered enterprise chatbot using Python and NLP techniques, automating log issue resolution and saving 15 hours/week across ticketing and testing departments
- Enhanced query accuracy by 25% through implementation of BGE3-Large vector embeddings, achieving 92% user satisfaction rate across 500+ weekly interactions
- Integrated real-time database retrieval with optimized caching mechanisms, reducing response times by 30% and supporting enterprise-scale deployment
- Selected to present project at Nokia Bangalore University Connect (NBUC) program to academic partners and senior technical leadership, demonstrating innovative automated support system

Projects

Agentic RAG System for Trading Knowledge **Source code**

- Engineered state-of-the-art Retrieval-Augmented Generation system achieving 97% accuracy in trading queries by implementing Llama-3-70B with vector databases and contextual retrieval
- Implemented contextual chunking using Claude AI that improved retrieval relevance by 42%, processing 10,000+ document chunks while maintaining sub-150ms query response times
- Developed GPU-accelerated document processing pipeline in Python handling EPUB and PDF formats, reducing embedding generation time by 78% compared to CPU-only processing
- Designed agentic architecture with LangChain that dynamically determines optimal retrieval strategies, delivering 3.5x more relevant results compared to traditional RAG systems
- Integrated multi-model flexibility supporting 3 LLMs (Llama-3, GPT-4o, Claude) with automated source citation, achieving 23% performance improvement through A/B testing

Autonomous Driving - Car Detection **Source code**

- Implemented YOLO object detection system for autonomous vehicles achieving 89% accuracy in identifying cars and traffic lights across various lighting conditions using TensorFlow
- Engineered critical algorithms including non-max suppression and IoU calculation in Python, eliminating redundant detections and improving localization precision by 35%
- Optimized tensor operations using TensorFlow to process 19×19×5×85 dimensional output volumes, enabling real-time detection capabilities at 30+ FPS
- Developed custom filtering algorithms in Python for extracting predictions from neural network outputs, generating accurate bounding boxes with 92% precision

Binary Image Classification with CNN **Source code**

- Developed robust image classification system using Convolutional Neural Networks in TensorFlow/Keras, distinguishing cat and dog images with 70% validation accuracy
- Implemented multi-layer CNN architecture with 32 and 64 filter convolutions using Python, achieving consistent performance improvement from 55% to 90% accuracy through hyperparameter optimization
- Engineered efficient data preprocessing pipeline in Python for normalizing 100×100×3 pixel images, optimizing training performance on 2,000+ samples
- Achieved model convergence through systematic architecture tuning and data augmentation techniques, reducing overfitting by 25%

Publications

From Nodes to Notables: A Graph Framework to Detect Emerging Influencers *2025, IEEE Conference*
Democratizing Machine Learning: A KNN-Guided Adaptive AutoML Framework *2025, IEEE Conference*
Smart Surveillance: AI-Driven Threat Detection and Women Safety Enhancement *2025, IEEE Conference*
Automated Q and A Chatbot: Harnessing AI for Efficient Information Retrieval *2025, SCOPUS (IRAJ)*

Technical Skills

Programming Languages: Python, Java, C++, SQL, R
AI/ML: LangChain, LlamaIndex, PyTorch, TensorFlow, Pandas, NumPy, RAG Systems, CrewAI
Web & APIs: FastAPI, Streamlit, REST APIs, OAuth, Redis
Tools & Infrastructure: Git, GitHub, Docker, PostgreSQL, AWS SQS
Visualization: Matplotlib, Seaborn, Plotly

Certificates

Deep Neural Networks with PyTorch **Coursera**
Certification Link

Deep Learning Specialization by Andrew Ng (5 Modules) **Coursera**
Specialization Link

Achievements

ETL Hackathon **2025**
Secured 1st Place *Dayananda Sagar University*

- Developed a data pipeline to extract cricket player statistics from Cricbuzz, transform the data for user comparison, and visualize results through interactive bar and pie charts.
- Ranked 1st out of 300 participating students, demonstrating exceptional data engineering and visualization skills.

Discord Bot Creation **2025**
Entertainment Project

- Engineered a Discord bot simulating Bitcoin mining, where users solve math problems ranging from simple to intermediate difficulty to earn virtual bitcoins.
- Enhanced user engagement through gamification, creating an educational tool that teaches both mathematical concepts and basic cryptocurrency principles.