

SWARAJ KHAN P

+91 8618893815 swarajkhan2003@gmail.com [LinkedIn](#) [Github](#) [LeetCode](#) [Portfolio](#) [YouTube](#)

A motivated Computer Science student with a strong passion for Python programming and machine learning. Seeking an internship to apply theoretical knowledge in real-world projects, enhance software development skills, and advance expertise in machine learning techniques.

Education

Dayananda Sagar University, Bengaluru <i>B.Tech in Computer Science & Engineering</i>	2021 - 2025 7/10 CGPA
The Amaatra Academy, Bengaluru <i>Class 12 PCMC</i>	2020 - 2021 79%

Work Experience

Draconic AI <i>Machine Learning Engineer — Founding Intern</i>	Oct '24 – Present <i>In-Office</i>
<ul style="list-style-type: none">Architected a comprehensive data engineering pipeline leveraging FastAPI endpoints to extract real-time OHLCV trading data from Zerodha and Dhan platforms, implementing dual-storage strategy with Redis for high-speed access and PostgreSQL (Supabase) for persistent storage.Developed advanced computation modules for critical trading metrics including candlestick patterns, price action indicators, swing analysis, and options chain analytics, optimizing data availability for algorithmic trading decisions.Designed and implemented a secure authentication system with Google OAuth and WhatsApp OTP verification, integrating with Supabase backend for robust user management in a production environment.Engineered multiple AI agent systems using CrewAI, Agno, and Smol frameworks, creating both Anthropic Claude-powered production agents and open-source HuggingFace alternatives, while implementing a comprehensive LLM monitoring system tracking token usage and response latency.Built dual-purpose Retrieval Augmented Generation (RAG) systems: (1) a general knowledge base indexing 60+ trading books by renowned authors, and (2) a personalized system enabling analysis of user-uploaded trading journals and TradingView charts combined with real-time market data.Conducted systematic prompt engineering experiments using PromptLayer to optimize LLM performance across different contexts, significantly enhancing response quality and reducing token consumption.	
Nokia <i>Chatbot Development — Intern</i>	Mar '24 – July '24 <i>Bangalore</i>
<ul style="list-style-type: none">Developed an AI-powered chatbot for Nokia's ticketing and testing teams, automating log issue resolution processes and saving approximately 15 hours per week in manual workload across both departments.Enhanced chatbot query accuracy by 25% through implementation of advanced NLP techniques and BGE3-Large vector embeddings, optimizing response relevance and achieving a 92% user satisfaction rate.Integrated real-time database retrieval with optimized caching mechanisms, cutting response times by 30% and improving scalability for enterprise-level deployment.Selected to present the project at Nokia Bangalore University Connect (NBUC) program, a prestigious event attended by academic partners and senior technical leadership, receiving recognition for innovative approach to automated support systems.	
Dexian <i>Machine Learning — Intern</i>	Jul '23 – Aug '23 <i>Bangalore</i>
<ul style="list-style-type: none">Designed and implemented a high-accuracy heart stroke prediction system using ensemble methods including Random Forest (94.5%), Gradient Boosting (94.1%), and Logistic Regression (75.4%), enabling early risk assessment for preventative care.Optimized data preprocessing pipeline with advanced feature engineering techniques, improving overall model performance by 10% while reducing computational overhead by 15%.Developed interactive visualization dashboards using Matplotlib and Seaborn to communicate critical insights to medical stakeholders, facilitating data-driven decision making for patient interventions.	

Projects

Binary Image Classification with CNN

 [Source code](#)

- Developed a robust image classification system using Convolutional Neural Networks (CNN) to distinguish between cat and dog images with 70% validation accuracy
- Implemented a multi-layer architecture with two convolutional layers (32 and 64 filters), max pooling, and dense connections using TensorFlow and Keras
- Engineered efficient data preprocessing pipeline for normalizing $100 \times 100 \times 3$ pixel images and optimizing model training on 2,000+ samples
- Achieved consistent performance improvement across training epochs from 55% to 90% accuracy through hyperparameter tuning and architecture optimization

Autonomous Driving - Car Detection

 [Source code](#)

- Implemented a state-of-the-art YOLO (You Only Look Once) object detection system for autonomous vehicles, achieving 89% accuracy in identifying cars, traffic lights, and other road objects under various lighting conditions
- Engineered critical algorithms including non-max suppression and intersection over union (IoU) calculation to eliminate redundant detections and improve localization precision
- Optimized tensor operations using TensorFlow to process the model's $19 \times 19 \times 5 \times 85$ dimensional output volume for real-time detection capabilities
- Developed custom filtering algorithms to extract meaningful predictions from complex neural network outputs, enabling accurate bounding box generation around detected objects

Agentic RAG System for Trading Knowledge

 [Source code](#)

- Engineered a state-of-the-art Retrieval-Augmented Generation system achieving 97% accuracy in answering trading queries, leveraging Llama-3-70B and vector databases for contextual information 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 a GPU-accelerated document processing pipeline that handles both EPUB and PDF formats, reducing embedding generation time by 78% compared to CPU-only processing
- Designed an agentic architecture with LangChain that dynamically determines optimal retrieval strategies, resulting in 3.5x more relevant results compared to traditional RAG systems
- Integrated multi-model flexibility supporting 3 different LLMs (Llama-3, GPT-4o, Claude) with automated source citation, enabling A/B testing that identified a 23% performance improvement using Llama-3-70B for financial content

Publications

Democratizing Machine Learning: A KNN-Guided Adaptive AutoML Framework

2025

Accepted for Publication

IEEE Conference

- Co-authored research introducing a novel AutoML pipeline that automates preprocessing, model selection, and hyperparameter tuning for structured and time-series data.
- Developed a K-Nearest Neighbors approach for neural architecture determination that achieved 91.19% average accuracy across ten CSV datasets and 63.66% loss reduction across time-series datasets.
- Created a flexible framework supporting multiple data types including CSV, time-series, and image datasets with specialized processing pipelines.

Smart Surveillance: AI-Driven Threat Detection and Women Safety Enhancement

2025

Accepted for Publication

IEEE Conference

- Collaborated on an innovative surveillance system that leverages Raspberry Pi 5 for edge computing, enabling real-time threat detection and response.
- Integrated multiple detection models including face emotion recognition, weapon detection, violence detection, and behavior analysis to create a comprehensive security solution.
- Developed the alert management system using Telegram chatbot for instant notification, allowing rapid response to potential security threats.
- Contributed to system evaluation achieving over 90% accuracy in violence detection and 86% accuracy in weapon detection across various test scenarios.

- Developed an AI-driven PDF-based Q&A chatbot utilizing text extraction, chunking, and cosine similarity matching for accurate information retrieval without relying on generative AI.
- Implemented advanced data sanitization techniques to prevent XSS attacks, enhancing security while maintaining sub-second response times.
- Created comprehensive logging mechanisms to track interactions and queries, improving system reliability and user experience.
- Presented research findings demonstrating how efficient document parsing and vector representation can outperform conventional NLP techniques for specific Q&A applications.

Technical Skills

AI/ML: LLM (Anthropic, Hugging Face), RAG Systems, Agentic Framework, PyTorch, Pandas, NumPy, Model Development, Feature Engineering
Languages: Python, SQL, C++, JAVA, R
Web Technologies: Chainlit, Streamlit, Web Scraping, Langchain, LlamaIndex
Developer Tools: Docker, Git, VS Code, GitHub
Databases & Infrastructure: Redis, PostgreSQL, Grafana, Logfire, Ray, SQS

Certificates

Deep Neural Networks with PyTorch	Coursera
<i>Certification Link</i>	
Deep Learning Specialization by Andrew Ng (5 Modules)	Coursera
<i>Specialization Link</i>	

Achievements

ETL Hackathon	2025
<i>Secured 1st Place</i>	
<i>Dayananda Sagar University</i>	
<ul style="list-style-type: none">• 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
<i>Entertainment Project</i>	
<ul style="list-style-type: none">• 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.	