Karthik Surya N A

+91-9360498834 | n.a.karthiksurya@gmail.com | linkedin.com/in/karthiksurya-na | github.com/nakarthiksurya

Summary

B.Tech in Artificial Intelligence and Data Science from Panimalar Institute of Technology with practical experience in building AI-driven systems. Currently working as a Software Engineer at C1Exchange (C1X), contributing to real-time intelligent agents, multi-agent frameworks, and retrieval-augmented generation (RAG) systems. Experienced in machine learning, natural language processing, and cloud-based model deployment. Skilled in Python, Transformers, FastAPI, Generative AI, CrewAI, and the Gemini API. Passionate about solving real-world problems with scalable AI solutions and committed to continuous learning in AI and software engineering.

Skills

Technical Skills

Programming & Scripting: Python, Typescript
Data Analysis & Visualization: Pandas, Matplotlib, Power BI
AI & Machine Learning: Generative AI, LLMs, BERT Models
Frameworks & Tools: Nest JS, FastAPI, Docker, Git, VS Code, WordPress, Bootstrap Studio
Platforms & Software: Windows, Linux (Ubuntu)

Soft Skills

Communication
Problem-solving
Teamwork and Collaboration
Active listening
Experimentation
Adaptability

Education

Panimalar Institute of Technology — Chennai, Tamil Nadu
Bachelor of Technology in Artificial Intelligence and Data Science,CGPA : 8.30 — *Nov. 2021 – May 2025*

Green Park International School(12th Grade) — Namakkal, Tamil Nadu
Higher Secondary School, Percentage: 77.4 — *2020 – 2021*

Green Park International School(10th Grade) — Namakkal, Tamil Nadu
Secondary School, Percentage: 77.4 — *2018 – 2019*

Experience

Software Engineer — July 2025 – Present
Class One Exchange(C1X) — *Chennai, India*

- Contributing to production-ready intelligent agents using Gemini APIs.
- Building end-to-end multi-agent systems integrating Slack, Jira, GitHub with Gemini 2.5 Flash.
- Developing and deployed Gradio-based RAG systems using ColBERT, CLIP, BM25 and Qdrant using NestJS Framework.
- Engineering in real-time contextual analytics and qualitative sentiment insights.

Internships

AI Intern – On-SIte — June 2025
Class One Exchange(C1X) — *Chennai, India*

- Built Slack sentiment chatbot using LLaMA 3.2B and Gemini API.
- Developed an MCP (Model Context Protocol) server for Google Gemini 2.5 integration.
- Worked on RAG-based agents using CLIP, ColBERT, BM25, and Qdrant for semantic search.

Data Science Intern – On-Site — July 2024 – September 2024
QBrainX — *Coimbatore, India*

- Increased project efficiency by 70% through developing and deploying a real-time analytics machine learning model.
- Utilized Python, HuggingFace Transformers and FastAPI for data preprocessing, model training, and deployment.
- Enhanced animal image classification accuracy and performance with pre-trained transformers.
- Developed an end-to-end Python application with FastAPI, including front-end integration.

Data Analytics Intern – On-Site June 2023 – July 2023
Juno Dynamics *Coimbatore, India*

- Developed and deployed cloud systems for SMEs, optimizing data analytics processes at a micro scale
- Gained hands-on experience in cloud systems, working on tools and technologies for data analytics in HAAS
- Collaborated with the team to enhance the efficiency and scalability of cloud-based solutions
- Exceeded performance expectations by delivering impactful solutions during the internship

Projects

LegalAdviser-AI (CivicAI) — Multi-Agent Assistant | *FastAPI, Gemini, Docker* Nov 2025

- Engineered an AI-powered legal assistant for Indian law, utilizing Gemini 2.5 Flash and RAG techniques to provide actionable guidance with validated case-law citations.
- Architected a multi-agent pipeline (Analyzer → Researcher → Summarizer) featuring automated intent detection to route queries effectively.
- Developed a targeted web-crawling module to ground responses in authoritative sources like IndianKanoon, reducing hallucinations.
- Implemented a robust backend using FastAPI with stateful session management and TTL-based automated cleanup.
- Containerized the application environment using Docker Compose, streamlining deployment with full OpenAPI/Swagger documentation.

Legal Information Retrieval System | *Python, Transformers, LlamaIndex, FastAPI* February 2025 – May 2025

- Developed a legal question-answering system using a custom-trained LegalBERT model integrated with Retrieval-Augmented Generation (RAG) techniques.
- Used semantic search and LlamaIndex to retrieve relevant Indian laws and case judgments based on user queries.
- Built a FastAPI-based backend and designed a user-friendly interface for public legal information access.
- Focused on the RTI Act and 49 other Indian laws with over 5000 case judgments for domain-specific training.

BeastlyVisionX:Animal Image Classifier | *Python, FastAPI, Docker* August 2024 – September 2024

- Developed an animal image classification system using HuggingFace Vision Transformer and FastAPI for real-time species recognition.
- Created a structured dataset with 90 attributes per class and deployed the model with Docker for scalability.
- Built a user-friendly front-end for seamless interaction with the classifier.

AI Story Generator | *Google's Gemini Pro, Python, NLP* March 2024

- Developed a story generation tool using Generative AI for real-time narrative creation.
- Assisted writers by generating unique, high-quality prompts and coherent stories.
- Received positive feedback for enhancing creativity and overcoming writer's block.

Publications

- **Towards Intelligent Legal Information Retrieval: A Transformer-Based Framework**
  Published in *International Journal of Scientific Research in Engineering Management (IJSREM)*, Volume 09, Issue 05, May 2025.
  DOI: 10.55041/IJSREM48260

Certifications

Microsoft Office Beginner and Advanced Course – Microsoft (Naan Mudhalvan), November 2022
Cybersecurity Analysis Design: – FutureSkills Prime (Gold Certified), June 2024
Introduction to Generative AI Learning Path: – Google Cloud (Coursera), March 2025
Transformer Models and BERT Model: - Google Cloud (Coursera), March 2025