

# Kanaka Raju Sampathi Rao

AIML Architect

+91 9640880425 | kanaka.raju95@gmail.com | linkedin.com/in/kanakarajusampathirao

## PROFESSIONAL SUMMARY

---

AI/ML Architect and Lead Data Scientist with over 10 years of experience delivering enterprise-grade Machine Learning, Generative AI, and multimodal video intelligence solutions. With a Master's in Computer Science from Georgia Institute of Technology, I bring strong foundations in AI, deep learning, and scalable systems. Proven expertise in Computer Vision, NLP, and Generative AI spanning LLMs, RAG pipelines, LangChain, and graph-based retrieval systems. Adept at architecting and deploying production-ready solutions across cloud, on-premise, and edge environments, including analytics agents, multimodal search/chatbots, and content automation platforms for Fortune 500 clients. Skilled in aligning AI architectures with business goals, integrating structured and unstructured data at scale, and leading cross-functional teams to deliver measurable impact.

## EXPERIENCE

---

### ClearTrail Technologies Pvt. Ltd.

*AIML Architect*

Hyderabad

*Dec 2025 – Present*

#### – Enterprise Video Search and Summarization Platform

- Architecting a large-scale video intelligence platform enabling semantic search, event discovery, and automated summarization across large video repositories.
- Designed an end-to-end multimodal pipeline combining computer vision, audio processing, and language models to extract spatio-temporal metadata from video streams.
- Implemented object detection, multi-object tracking, scene understanding, and speech-to-text pipelines to generate rich, queryable video representations.
- Enabled natural language querying over video data using embedding-based retrieval and LLM-driven reasoning for context-aware responses and timestamped insights.
- Optimized high-throughput inference pipelines using GPU-accelerated, containerized deployments for scalable and low-latency video analytics.
- Collaborating with platform and product teams to harden the system for enterprise-grade deployment in secure, on-premise and air-gapped environments.

### Coforge

*Lead Data Scientist*

Hyderabad

*Dec 2019 – Present*

#### – GraphRAG-Based Chatbot for Web Analytics Insights

- Developed an advanced RAG-powered chatbot to answer complex queries based on 100 internal presentation decks containing KPIs like bounce rate, scroll depth, and user flow across homepage, product, and family pages.
- Modeled a Neo4j knowledge graph where nodes represent products, geographies, metrics, and slide content (text, charts, tables), and edges capture semantic relationships.
- Embedded slide content into a vector database and implemented semantic retrieval + graph traversal for context-grounded response generation via LLMs.
- LLM-Based Content Automation for Copywriting
- Automated generation of compliant marketing content using a two-stage GenAI pipeline.
- Deployed Gemini Enterprise API for editorially consistent text variant generation and Pixtral LLM (hosted on a secure air-gapped AWS instance) for generating accessibility-compliant Alt Text for product images.
- Implemented RAG-based few-shot prompting and agentic flows to enforce tone, grammar, product accuracy, and length constraints.
- Automatic Git PR Review Assistant
- Built an intelligent assistant that generates Git PR comments using prompt-engineered LLMs to reflect developer, reviewer, and QA engineer personas.
- Enabled auto-summarization of PR diffs and code context to streamline code reviews, enforce guidelines, and improve engineering feedback loops.
- Enterprise GenAI SQL Agent
- Developed an AI-powered SQL Agent that translates natural language questions into SQL queries and executes them on enterprise databases.

- Enabled automated result retrieval and generation of interactive visualizations, enhancing data exploration and decision-making.
- Implemented context-aware query generation with dynamic summarization of wide tables for improved accuracy and usability.
- Boosted analyst productivity by automating query execution, visualization, and KPI extraction from large-scale datasets.
- **Active Learning-Based Annotation Platform (Zastra)**
- Designed and deployed an end-to-end platform that validates annotation quality using active learning loops and model-in-the-loop architecture.
- Reduced manual QA effort and boosted model training efficiency by enabling iterative, confidence-aware validation cycles.
- **Detecting CVI Disease Progression**
- Built a medical imaging solution to detect and quantify morphological changes in bones and muscles of diseased legs for Chronic Venous Insufficiency (CVI).
- Used image-based measurements to assess disease progression and assist clinical diagnostics.
- **Digital Fingerprinting for Coil Tracking**
- Developed a computer vision-based tracking system using IP cameras to capture images across manufacturing checkpoints (Autobrazzer to Helium Leak Test).
- Extracted unique digital fingerprints from images to trace coil identity, enabling end-to-end quality control.
- **Multi-Label Product Listing for E-Commerce**
- Developed a multi-label classification model to automate the generation of product listing attributes, minimizing manual tagging effort.
- Improved listing accuracy and consistency for large-scale e-commerce catalog ingestion pipelines.

## DigiTele Networks Pvt Ltd (Greenko Group)

Hyderabad

*Data Scientist*

*Aug 2018 – Nov 2019*

- Developed drone-based image analysis pipeline to detect defective solar panels and wind turbine blades using IR and RGB imagery.
- Automated detection of anomalies like dirt, shade, glass breakage, laminate damage, and blade detachment.
- Built an ALPR system using image processing and OCR on RTSP camera streams to extract real-time vehicle registration data.

## ValueLabs

Hyderabad

*Senior Software Engineer*

*Jan 2016 – Jul 2018*

- Built an AI-powered pipeline to assess automobile panel collision damage severity and recommend repair or replacement.
- Developed a Django-based web application to extract unstructured data from Word documents and convert it to structured XML.

## EDUCATION

---

### Georgia Institute of Technology

Atlanta, GA

*M.S. in Computer Science (OMSCS)*

*Jan 2023 – Aug 2025*

### JNTU Hyderabad

Hyderabad, India

*B.Tech in Mechanical Engineering*

*Jul 2011 – May 2015*

## SKILLS

---

**Programming Languages & Web Frameworks:** Python, Flask, Django

**Generative AI & LLM Ecosystem:** LangChain, ChromaDB, Pinecone, Milvus, OpenAI, vLLM, Pixtral, Gemini, Ollama, LLM Agents, RAG, Prompt Engineering, ReAct Agents

**Machine Learning & Computer Vision:** TensorFlow, PyTorch, Scikit-learn, OpenCV, DoWhy, Pandas

**Cloud & Infrastructure Tools:** AWS, Azure, Docker, Linux

**Development Tools & CI/CD:** Git, GitHub Webhooks, Postman, VSCode

**Databases & Graph Technologies:** SQL, MongoDB, Neo4j

**Model Deployment & Serving:** Flask APIs, FastAPI, Dockerized Deployments, GPU-Based Model Serving