

# Mohan Krishna G R

mohankrishnagr08@gmail.com | +91-9080696129 | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

## EDUCATION

### Sri Ramakrishna Engineering College

Master of Technology, Computer Science & Engineering (5 Years Integrated)

CGPA: 9.17/10.0 | Relevant Coursework: Deep Learning, Data Structures & Algorithms, Distributed Systems

Coimbatore, India  
Sept 2022 – March 2027

## RESEARCH & EXPERIENCE

### AI Engineer Intern | 2Cents Capital

Focus: Multi-Agent Systems and Asset Management Research

- Engineering **Agentic AI systems** (Valura.ai) for investment research and asset-management decisioning.

Dubai, UAE (Remote)  
Jan 2026 – Feb 2026

### Generative AI Intern | AIRBUS

Research focus: GraphRAG, Explainable AI, and Transformer Optimization

- Investigated **graph-augmented retrieval methods** for long-context LLM reasoning on complex engineering corpora; proposed a GraphRAG variant achieving **25% higher retrieval accuracy** over dense-only baselines.
- Designed a **causal graph-conditioned RAG framework** enabling interpretable root-cause analysis in manufacturing workflows; validated explanations via **counterfactual perturbation studies** and expert review - **presented** at **Airbus Global AI Week 2025**.
- Conducted **transformer inference optimization** using quantization and PEFT, achieving a **2× throughput improvement** with negligible accuracy degradation for production deployment.
- Authored a research paper on **agentic document extraction** for engineering stress dossiers; **received** a **Top 5 Paper Award** among 350+ engineers at **Airbus TechXceed 2025**.

Bengaluru, India  
Jun 2025 – Dec 2025

### AI/ML Intern | Infosys

Focus: Natural Language Processing & Transformer Fine-tuning

- TextSumm**: Large-Scale Summarization Study. Conducted an empirical comparison of **extractive vs. abstractive summarization** on a custom **550K+ document corpus**, evaluating performance across ROUGE metrics.
- Achieved a **3× ROUGE-2 improvement over abstractive baselines** using fine-tuned transformer models combined with lightweight statistical heuristics.

Remote  
May 2024 – Jul 2024

## PUBLICATIONS

- Indian Rainfall Prediction Using Machine Learning: A Comparative Study | [IEEE Xplore](#), [ICONSCEPT 2024](#) - NIT Py.

## RESEARCH PROJECTS

### CertusHire — Multi-Agent LLM Evaluation System

Individual Project

GitHub: [Main Repository](#) | [Multi-Agent Coordinator](#) | [RAG Retrieval Engine](#) | [Voice-Enabled Candidate UI](#) | [Recruiter Analytics](#) | [DAR3 Protocol](#)

- Investigated **multi-agent LLM coordination** for adaptive technical evaluation; designed a **state-machine-based orchestration framework** enabling dynamic role assignment between interviewer and evaluator agents.
- Proposed a **hybrid dense-sparse RAG pipeline** to ground generation in verified sources, achieving **85% agreement with human interviewer judgments**.
- Evaluated scalability and latency trade-offs of real-time multi-agent inference, sustaining **<200ms response latency** for **100+ concurrent sessions** under fault-tolerant scheduling.
- Tech Stack**: Python, FastAPI, React, TypeScript, PostgreSQL, Celery, RabbitMQ, Docker, LangChain, llama.cpp, RAG

### PyroGuardian — Edge-AI Computer Vision System

Team Project

GitHub: [Main Repository](#) | [Edge AI Engine](#) | [Cloud Dashboard](#) | [Research Baseline](#)

- Investigated **efficient deep learning inference under edge constraints** for real-time fire detection; optimized an **86.7M-parameter model** for deployment on NVIDIA Jetson Nano.
- Applied **TensorRT FP16 optimization** and CUDA acceleration, achieving a **90% inference speed-up** over PyTorch FP32 while sustaining **30 FPS at 720p**.
- Constructed and curated a **51GB multi-scenario dataset** (10K+ annotated frames across 8 fire conditions); improved robustness via targeted augmentations and evaluated against YOLO-based baselines.
- Tech Stack**: Python, OpenCV, PyTorch, RT-DETR, TensorRT, CUDA, NVIDIA TAO, AWS SNS, DeepStream, Nvidia Jetson Nano.

## TECHNICAL SKILLS

- AI/ML & Research**: Machine Learning, Deep Learning, Transformers, Representation Learning, Retrieval-Augmented Generation, Multi-Agent Systems, Model Evaluation & Analysis.
- NLP & LLMs**: Sequence-to-Sequence Models, Embeddings, Dense & Sparse Retrieval, Hybrid RAG, Prompting, Fine-Tuning (PEFT / LoRA), Preference Optimization (DPO, GRPO).
- Efficient ML & Systems**: Inference Optimization, Quantization, Latency-Throughput Trade-offs, Edge Deployment, Model Serving.
- Frameworks & Libraries**: PyTorch, JAX, TensorFlow, Hugging Face, CUDA, TensorRT.
- Languages**: Python, C/C++, Java, SQL.

## HONORS & LEADERSHIP

- Sri. P. Ramaswamy Memorial Award**: For academic excellence — **'The Highest CGPA'** all 3 years (Top 0.3% of 4,400+ students).
- Runner-Up (Honeywell Drone Hackathon)**: Built PyroGuardian UAV fire-detection system using TensorRT. (Among 200+ teams)
- Overall Winner (The Ultimate Hackathon – CII)**: Developed **MindWave**, an AI mental-health app (94% accuracy). (Among 1000+ teams)