

Mohan Krishna G R



mohankrishnagro8@gmail.com



+91-9080696129



[Portfolio](#)



[LinkedIn](#)



[GitHub](#)

Education

Sri Ramakrishna Engineering College

September 2022 – March 2027

Master of Technology, Computer Science & Engineering (5 Years Integrated) | **CGPA: 9.17/10**

Coimbatore, India

Relevant Coursework: Data Structures & Algorithms, Operating System, DBMS, Distributed Systems, Computer Networks

Experience

AI Engineer Intern @ [2Cents Capital](#)

January 2026 – Present

Internship

Dubai, UAE (Remote)

- Engineering Agentic AI systems for investment research and asset-management decisioning at Valura.ai.

Generative AI Intern @ [AIRBUS](#) | [LinkedIn Recommendation](#)

June 2025 – December 2025

Internship

Bengaluru, IN

- 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.
- 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**.
- 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**.
- Conducted **transformer inference optimization** using quantization and PEFT, achieving a 2x throughput improvement with negligible accuracy degradation for production deployment.

Trainee Software Engineer @ [SuperDNA 3D Lab](#)

January 2025 – March 2025

Part-Time

Remote

- Developed **CV pipelines** for automated QA of AI-generated e-commerce imagery (15-rule rubric), **dimension estimation in monocular image** using depth transformers, and experimental **3D hair reconstruction** (UniHair).

AI/ML Intern @ [Infosys Springboard](#) | [LinkedIn Recommendation](#)

May 2024 – July 2024

Summer Internship

Remote

- Developed a research-grade **text summarization** pipeline on 564K+ records with **transformer fine-tuning** (300% ROUGE-2 improvement) and novel **TF-IDF + KMeans** modelling; deployed reproducible experiments on **Azure**.

Publications

Indian Rainfall Prediction Using Machine Learning: A Comparative Study | [IEEE Xplore](#) | [DOI](#) | *IConSCEPT 2024 - NIT Py*

Hackathons

Honeywell Drone Technologies Hackathon 2024 | [Runner-Up](#) | ₹25,000 (~\$300)

- **Secured 2nd place** among 200+ teams for developing **PyroGuardian**, a UAV fire-detection system using **TensorRT-optimized models** (90% faster inference) and **CUDA-accelerated edge deployment** on NVIDIA Jetson Nano.

The Ultimate Hackathon 2023 (CII) | [Overall Winner](#) | ₹25,000 (~\$300)

- **Won 1st place among 1000+ teams** for **MindWave**, a real-time **AI mental-health app** (94% accuracy) featuring optimized low-latency APIs and scalable, analytics-driven architecture.

Projects

CertusHire - Autonomous Multi-Agent Interview Platform | [GitHub](#)

Individual Project

Tech Stack: Python, FastAPI, React, TypeScript, PostgreSQL, Celery, RabbitMQ, Docker, LangChain, llama.cpp, RAG

- Built a **production-grade autonomous multi-agent LLM system** for end-to-end technical interviews (JD/CV parsing, adaptive voice questioning, automated evaluation), reducing manual screening effort by **90%**.
- Designed **DAR3**, a **state-machine-based multi AI Agents orchestration framework** coordinating interviewer and evaluator agents to dynamically adjust question difficulty based on real-time candidate performance.
- Implemented **idempotent APIs**, queue-backed task execution, and failure recovery using Celery/RabbitMQ to ensure consistency under partial failures and concurrent requests.
- Implemented a **hybrid dense-sparse RAG pipeline** (embeddings + TF-IDF) to ground generation in verified sources, achieving **85% agreement with human interviewer scores**.

- Scaled real-time conversational inference to **100+ concurrent sessions** with a **fault-tolerant FastAPI + Celery/RabbitMQ backend**, maintaining **<200ms latency**, and delivered a low-latency React/WebSocket frontend **with Constitutional AI safety guardrails**.

TextSumm - Document Summarization Platform | [GitHub](#) | [DockerHub](#)

Individual Intern

Tech Stack: Python, FastAPI, Docker, GitHub Actions, CI/CD, Azure, AWS, PyTorch, Hugging Face Transformers, Scikit-learn, NLP, REST APIs, HTML5, CSS3, JavaScript.

- Built an **end-to-end text-summarization** product that converts long documents (PDF, DOCX, URLs, raw text) into concise, high-quality summaries using both abstractive and extractive NLP pipelines.
- **Boosted summarization accuracy by 300%** (ROUGE-2 F1-score) via hybrid NLP architecture: fine-tuned transformer models (ROUGE-1: 61.32) [abstractive] + novel rule-based extractive model (TF-IDF/KMeans) on **550K+ records**.
- Engineered multi-format REST API (FastAPI) supporting PDFs, DOCX, and URLs, paired with a responsive frontend, **reducing user processing time by 40%**.
- **Deployed at scale** via **Docker** on **Azure ACI** with automated CI/CD (**GitHub Actions**), achieving **99.9% uptime** and seamless updates.

PyroGuardian - Real-Time Fire Detection System | (Source Code & Models with [Honeywell](#))

Team Project

Tech Stack: Python, OpenCV, PyTorch, RT-DETR, TensorRT, CUDA, NVIDIA TAO, AWS SNS, DeepStream, Nvidia Jetson Nano.

- **Led a multi-disciplinary team** to build an AI-powered UAV (Drone) system, **reducing fire response time by 40% vs. thermal sensors** and achieving **20% fewer false alarms than traditional video analytics**.
- **Deployed an 86.7M-parameter model** on Jetson Nano (TensorRT-optimized FP16), achieving **90% faster inference vs. PyTorch FP32** and **30 FPS at 720p** for real-time severity analysis.
- **Engineered 51GB dataset** (10K+ annotated frames across 8 fire scenarios) with augmentations (cropping/colour jittering) for multi-scale robustness.
- **Integrated AWS SNS alerts** with RBAC roles, achieving **<1s latency** for 500+ users and dynamic prioritization via severity scores.
- **Deployed in 3+ industries** with **low-bandwidth streaming** (OpenCV/H.264), reducing video feed downtime by **20%** in field trials.

Crop Schedule Management (Quantum) | (Guided by SREC & CQuICC, IIT Madras) | [GitHub](#)

Team Project

Tech Stack: Python, D-Wave Ocean SDK, dimod, PyYAML, Click, Matplotlib, Quantum Annealing (D-Wave Leap Hybrid Solver).

- **Reduced computational overhead by 50%** vs. classical solvers by formulating crop scheduling as a **QUBO model** and solving via D-Wave's quantum annealing. **Automated optimal crop rotations** for 10+ crop types across 100+ fields, adhering to adjacency rules and resource constraints (water/fertilizer).
- **Achieved 95% field utilization** in simulations for Indian agriculture, improving yield predictions by 30% over traditional methods.

MindWave - AI Mental Health Platform | [GitHub](#)

Team Project

Tech Stack: Flutter, Dart, Python, Flask, REST APIs, HTTP Networking, Firebase (Auth, Firestore, Cloud Messaging), TensorFlow, Scikit-learn.

- **Led a 5-person team** to develop an AI-powered mental health app, achieving **94% accuracy** in stress prediction (validated via 5-fold CV) using a hybrid ML model (TensorFlow/scikit-learn).
- **Deployed model with TensorFlow Serving** and optimized API latency by **20%** via asynchronous processing, reducing server response time to <500ms for concurrent users.
- **Integrated real-time inference** into a Flutter mobile app (Android/iOS), enabling dynamic stress assessments and boosting user engagement by **35%** (tracked via Firebase Analytics).

Awards / Extra-Curricular

- Sri. P. Ramaswamy Memorial Award | **Highest CGPA** | A.Y. 2022-2025 | [Awarded](#) for academic excellence (top 0.3% of 4,400+ students).
- **Student Innovation Ambassador** | [Selected](#) 7/4,400 | Spearheaded campus tech initiatives, mentoring and inspiring 50+ students on AI/ML projects.
- **Best Project Award** | A.Y. 2023-2024 | [Recognized](#) for **Crop Schedule Management (Quantum)**, improving agricultural efficiency by **30%** through quantum optimization (QUBO/D-Wave).
- **Best Innovation Idea** | [InnoTech](#) 2023 | AI/IoT agricultural platform (UAV-based pest detection, IoT soil analytics).
- **Head, Blog Writers** | SREC | Led 80+ member team to publish technical articles, myself on AI/ML, Engineering, Project-based Learning, boosting campus engagement by 40%.

Technical Skills

Areas of Interest: Full Stack Development, System Design, Competitive Programming, AI/ML Systems.

Languages and Frameworks: C, C++, Python, Java, MySQL, SQLite, PostgreSQL, JS, Dart, HTML5, CSS3, Flask, React.js, Android, Node.js, Flutter, TailwindCSS, Express.js, Material 3, Bootstrap, Next.js, Haskell, Solidity, Vue.js, Django, AngularJS, SCSS, Go, gRPC

Tools and Technologies: Git, Linux, Firebase, Isar, Figma, RaspberryPi, Eclipse, Visual Studio, Arduino, AWS EC2, Azure VM, Remix IDE