

KOLLI KRUPA JOHNSON CHOWDARY

Full Stack AI Engineer · Generative AI · RAG · AI Agents
Hyderabad, India | kollikrupajohnson@gmail.com | [GitHub](#) | [LinkedIn](#)

SUMMARY

Full stack AI engineer (B.Tech CSE, Data Science specialization, CGPA 9.1 out of 10) who builds GenAI applications end to end, from React and Flutter frontends to FastAPI backends, RAG pipelines, and AI agents. Shipped 6 production systems as a software engineering intern, including a platform delivered to a sister company and an LLM observability stack, and lead the database design for BeverageOS. Built Zonny, an AutoGen multi agent platform with a 17 tool registry, ChromaDB RAG memory, and an MCP server, plus a one-command deploy CLI published on PyPI. AWS SAA-C03 and Azure AZ-900 certified.

SKILLS

Languages	Python, TypeScript, JavaScript, SQL, Bash, Dart, C++, HTML5, CSS3
GenAI and Conversational AI	LLMs, prompt engineering, RAG pipelines, embeddings, AI agents, tool calling, MCP
AI Frameworks	LangChain, LangGraph, AutoGen, Ollama, Groq API
ML and Vision	PyTorch, YOLOv8, TensorFlow Lite, OpenCV, scikit-learn, federated learning
Frontend and Mobile	React.js, Next.js, Flutter, Streamlit, HTML5, CSS3
Backend and APIs	FastAPI, Flask, NestJS, Node.js, REST APIs, GraphQL, WebSockets, RBAC
Vector DBs and Data	Qdrant, ChromaDB, PostgreSQL, Supabase, MySQL, MongoDB, Redis, TimescaleDB, Memgraph, PostGIS
Cloud and DevOps	AWS, Azure, Docker, Kubernetes, GitHub Actions CI/CD, Nginx, Linux, Git

WORK EXPERIENCE

Software Engineer / GenAI Developer Intern · Systems Group	Oct 2025 – Present Hyderabad, India
BeverageOS (Tech Lead): Designed the PostgreSQL schema for a beverage POS and kitchen display platform, covering modifier groups, ingredient ownership, and kitchen-display JSON; modeled inventory and recipes with 245 real supplier items and led the move to a NestJS modular monolith. WorkStream Platform: Delivered 1 internal platform to a sister company that replaced 3 disconnected tools, covering task management, team hierarchy, document versioning, and messaging. LLM Observability Stack: Improved LLM inference reliability by instrumenting 3 production metrics with structured logging and live health dashboards for monitoring and troubleshooting. Pixel Streaming Platform: Cut project onboarding from hours to minutes by engineering a 4 stage provisioning pipeline with 3 tier RBAC and session queue management. LMS Teacher Dashboard: Removed administrative bottlenecks across 3 educator workflows by centralizing class management, assignment tracking, and quiz assessment. CAD Comparison Tool: Eliminated manual drawing review across 3 change types by building a 4 stage pipeline covering parsing, diffing, visualization, and PDF reporting. Distro Streaming Ecosystem: Enabled secure kiosk streaming across a 2 component architecture using a Flutter Android client and virtualized server backend.	

PROJECTS

Zonny: Multi Agent AI Platform and Developer CLI	2024 – 2025
Tools: Python, AutoGen, LangChain, Ollama, ChromaDB, Redis, MCP, tree-sitter, FastAPI, Docker, Git, PyPI	
Multi agent orchestration: Coordinated a 6 agent AutoGen pipeline over a 17 tool registry, with 3 tier permissions controlling which tools each agent can call, and an auto-repair loop with up to 3 fix attempts. RAG based agent memory: Removed context loss between agent runs by building a retrieval layer over ChromaDB embeddings with Redis session caching. Testing and integration: Validated platform health with a 35 check diagnostic suite and exposed agent tools to external clients through an MCP server. ZonnyHelper CLI: Reduced deployment to a single command across 12 targets by publishing a PyPI CLI with tree-sitter code parsing, Git automation, and fallback across 4 LLM providers.	
LLM Complaint Management Agent	2026
Tools: Python, LangGraph, Groq API	
Conversational agent workflow: Built a complaint management system as a LangGraph state graph backed by Groq hosted LLMs for an AI engineering take home assignment.	

Geospatial Intelligence Foundry: Self Extending Ontology Engine

2024 – 2025

Tools: Python, Memgraph, TimescaleDB, PostGIS, Qdrant, Redis, MinIO, Docker

- **LLM assisted classification:** Attained full audit traceability across 4 decision types via a 3 tier pipeline of deterministic rules, fine tuned ML, and async LLM escalation.
- **Multi store data architecture:** Reduced query latency on high throughput workloads by designing a 5 store layer, including Qdrant vector search, each store selected for its access pattern.

GenAI Data Visualization Tool

Tools: Python, Streamlit, Plotly, spaCy, Transformers

- **Natural language analytics:** Turned plain-English questions about an uploaded CSV into interactive charts, with data profiling and quality checks built in.

Agricultural Plant Disease Detection System (Volunteer Team Project for PJTAU)

2024

Tools: Python, Flask, Flutter

- **End to end diagnosis app:** Enabled crop disease and pest identification for farmers through a Flutter photo capture app paired with a Flask admin review dashboard.

FedFusion: Federated Learning with Differential Privacy

2024

Tools: Python, PyTorch, YOLOv8, FedAvg

- **Privacy preserving model training:** Sustained convergence across non IID distributions by fine tuning YOLOv8 on distributed nodes using FedAvg with gradient clipping and noise injection.

Autonomous Vision System

2024

Tools: TensorFlow Lite, MobileNet SSD, OpenCV, Raspberry Pi

- **On-device detection:** Ran real-time object detection on a Raspberry Pi 4 with no cloud, driving motors directly from what the camera sees.

Personal Cloud Homelab

2025 – Present

Tools: Linux, Docker, Nginx, PostgreSQL, MinIO, Portainer, SSH

- **Self operated production infrastructure:** Preserved continuous uptime across 4 containerized services by configuring Nginx reverse proxy on a fully SSH managed Linux server.

Line Following Robot

Tools: Arduino (C++), Python

- **Embedded control:** Built an Arduino robot with 3 IR sensors that remembers its last turn to cross gaps and spins in place on sharp corners.

EDUCATION

B.Tech, Computer Science and Engineering with Data Science Specialization

2022 – 2026 | Hyderabad, India

· KLEF Deemed University

- CGPA: 9.1 out of 10.0 · Graduated 2026
- Coursework: Machine Learning, Deep Learning, Computer Vision, Cloud Computing, Distributed Systems, Data Structures and Algorithms, Database Management, Operating Systems.

CERTIFICATIONS

- AWS Certified Solutions Architect, Associate SAA-C03 · Microsoft Azure Fundamentals AZ-900 · AWS Academy
- Salesforce Agentforce AI Specialist AI-201 · GitHub Foundations · Red Hat Certified Enterprise Java EX183
- MongoDB Associate Developer · J.P. Morgan SWE Job Simulation · LinkedIn Communications Foundation

LANGUAGES

- English: Professional working proficiency · Telugu: Native proficiency · German: Beginner