G Sudarshan Sastry

AI Engineer | LLM Agents · RAG · Automation

siddhusastry333@gmail.com +91 76750 53840 Hyderabad, India sudarshan.krait.co.in
linkedin.com/in/sudarshansastry github.com/sudarshan032

Summary

AI Engineer with 1+ year of industry experience building LLM, RAG and agent systems end to end, from FastAPI services and vector retrieval to n8n automations and Dockerised cloud deployments. At Freedom With AI I delivered 14+ LLM/RAG systems for 9 client accounts and moved from AI Intern to Associate AI Developer. First-author IEEE IGARSS 2025 paper on deep reinforcement learning. Available immediately.

Skills

AI & LLM LLM agents, RAG, LangGraph, LangChain, MCP, tool calling, prompt engineering, LLM evals, embeddings and vector databases, Claude / OpenAI / Gemini APIs, Ollama, PyTorch

Automation n8n, webhooks, Playwright, workflow automation

Backend & Web Python, FastAPI, Django, Flask, Node.js, TypeScript, Next.js, React, REST APIs, WebSockets

Data & Cloud PostgreSQL, Redis, MongoDB, Docker, Linux, AWS, GCP, Vercel, GitHub Actions

Languages Python, Java, C, SQL

Experience

Freedom With AI Remote
Associate AI Developer (Full-time) Apr 2026 – Jul 2026
AI Intern Nov 2025 – Mar 2026

- Delivered 14+ LLM and RAG systems for 9 client accounts, from requirements to deployment, by building LLM routing with fallback, vector retrieval, FastAPI and Next.js services and n8n workflows, run as 16+ Dockerised services on AWS, GCP and OCI.
- Built a webinar and call attribution platform (Next.js, PostgreSQL) used by 5 client tenants by merging 4 export sources into one schema, and measured an AI calling channel across 100+ webinars: calls lifted purchases by 20–70% vs 4–5% for attendance.

Deep Algorithms Solutions Pvt Ltd Hyderabad, India
Software Engineer May 2025 – Oct 2025

- Built a DDoS detection pipeline that aggregates packets into flows and scores per-source packet rate, source-IP entropy and protocol mix against tunable thresholds.
- Added TLS fingerprinting from ClientHello metadata so detections key on the client's software stack rather than its IP; wrote the feature specs and acceptance tests with QA.

CAIR Lab, IIT Mandi Dec 2024 – Mar 2025
Research Intern

- Wrote STM32 firmware in C for BLDC motor control (timer-driven PWM commutation, closed-loop speed control), from bare-metal bring-up to bench tests.
- Fused readings from multiple gas sensors on Arduino with per-sensor calibration curves, drift compensation and noise filtering.

AKOSCOWL DOT INVENTIONS Pvt Ltd Hyderabad, India
AI-ML Intern Feb 2024 – Sep 2024

- Designed drone power and control electronics in KiCad and modelled the airframe in ROS/Gazebo with sensor plugins.
- Built obstacle-avoidance planners that send MAVLink waypoints through QGroundControl missions, validated in simulation before flight tests.

Earlier internships (8): Python, ML and automation Jul 2022 – Mar 2024
CodeClause · Bharat Intern · Coincent · Coders Cave · AICTE–EduSkills with AWS Academy (2), Google for Developers and SS&C Blue Prism

Projects

Change Guardian Oct 2026

Python · LangGraph · FalkorDB · MCP · FastAPI · Docker

- Exposed the real blast radius of IT changes (a "low-risk" patch that actually reaches 51 hospitals) by building a LangGraph agent that compares what the CMDB declares with what traffic, scans and logins prove, through 18 read-only graph tools served over MCP.
- Showed FalkorDB runs blast-radius traversals ~9x faster than PostgreSQL at 3.47M nodes (38 ms vs 357 ms; p95 1.7 s vs 4.4 s at 100 concurrent clients) with a reproducible benchmark, backed by 128 automated tests.

AppFlowy-AI Jan 2026

Flutter · Rust · FastAPI · LangChain · Ollama · Qdrant

- Turned plain-language requests into AppFlowy databases and properties fully offline by building an agent router that dispatches to a local Ollama model (gpt-oss:20b) through FastAPI and LangChain services.
- Kept the app and the agent in sync by rendering every database and record as Markdown with YAML front matter and watching file edits for two-way updates, with Qdrant search over records; packaged with Docker and nginx.

Claude Booths Oct 2026

Go · tmux · git worktrees · GitHub Actions

- Let one developer run several AI coding agents at once without edit conflicts by giving each task its own git worktree, branch and tmux session in a Go terminal app (hard fork of claude-squad).
- Made sessions reliable and releases verifiable by streaming tmux pane changes instead of polling, getting agents past trust and first-run screens automatically, and shipping signed-provenance releases with a checksum-verified installer.

UAV Obstacle Avoidance with Deep RL 2024

Python · DQL · DDPG · Unity3D

- Compared 2D LiDAR with camera depth estimation for fixed-wing UAV obstacle avoidance by training DQL and DDPG agents in a Unity3D simulator driven by a Python RL loop (B.Tech project).
- Both agents reached the goal collision-free after 300k+ training steps; LiDAR earned higher reward while depth estimation learned faster. Published as a first-author paper at IEEE IGARSS 2025.

Education

KL University (Koneru Lakshmaiah Education Foundation)

B.Tech, Electronics and Communication Engineering • CGPA 8.59/10

Specialization: Data Science & Big Data Analytics Hyderabad, India • 2021 – 2025

Publications & Awards

Obstacle Avoidance in Drones using Deep Reinforcement Learning 2025

IEEE IGARSS 2025 · first-author paper

1st Place, JHub 2.0 Regional Hackathon Jun 2022

JNTUH J-Hub

Amazon ML Summer School (selected via online assessment) 2023

Amazon