

KURMA LAKSHMI NARASIMHA VINESH

+91-9398948695 | vinesharceus@gmail.com | linkedin.com/in/vinesh007
github.com/VineshF1 | vinesh007.vercel.app

Summary

AI engineer and B.Tech (AIML) student specializing in multi-agent systems, Retrieval-Augmented Generation (RAG) pipelines, and LLM fine-tuning. Skilled in Python, LangChain, FastAPI, Google Agent Development Kit (ADK), Model Context Protocol (MCP), and vector databases (PostgreSQL, pgvector, Supabase). Experienced in building agentic AI workflows and shipping full-stack AI applications deployed with Docker and Vercel.

Skills

Programming Languages: Python, JavaScript

AI/ML: RAG, Multi-Agent Systems, Agentic AI, LLM Fine-tuning (QLoRA/Unsloth), MCP, Google ADK, NLP

Frameworks: FastAPI, LangChain, Unsloth

Databases & Vector Search: PostgreSQL, Supabase, pgvector

Developer Tools: Git, GitHub, Docker, Vercel

Projects

SkyTrace – Multi-Agent Satellite Tracking System — Python, Google ADK, MCP 2026

- Architected a multi-agent satellite tracking system using Google ADK, orchestrating Notification and Telemetry agents via a Model Context Protocol (MCP) gateway, demonstrating applied problem-solving with autonomous AI systems.
- Integrated Celestrak, N2YO, and Nominatim REST APIs with SGP4 propagation, ECEF distance computation, and token-bucket rate limiting to ensure accurate, real-time telemetry processing.

Halos AI – Formula 1 AI Assistant — Python, FastAPI, PostgreSQL + pgvector 2026

- Built a RAG + Text-to-SQL chatbot for F1 trivia and statistics using FastAPI and Supabase, applying NLP techniques and generative AI models (GPT-class LLMs) for natural language understanding.
- Achieved intelligent query routing via semantic search with NVIDIA NV-Embed-v1 embeddings, combining unsupervised vector retrieval with structured SQL generation on race data from 2021 to 2026.
- Engineered an F1 knowledge base from Wikipedia with data preprocessing and feature engineering, achieving high factual grounding for retrieval-augmented generation.

EauRouge-F1 – Qwen3-8B Fine-tuning — Python, Unsloth, QLoRA 2026

- Curated a 10,656-example Formula 1 QA dataset by aggregating and normalizing 1,149 races (75 seasons, 1950–2025) from the Jolpica API and Wikipedia into paraphrased QA pairs across 11 categories.
- Fine-tuned Qwen3-8B (4-bit quantized) via QLoRA (rank-32 LoRA adapters, Unsloth, Colab T4) with cosine LR scheduling and chat-template prompting; published weights to Hugging Face.

Certifications

- **Introduction to Python** – GUVI (2025): Core Python — syntax, data structures, loops, functions — applied to practical problems.
- **Oracle Agentic AI Foundations Associate** – Oracle (2026): Autonomous AI agents, from orchestration fundamentals to real-world agentic design on Oracle's cloud platform.
- **Hugging Face Agents Course** – Hugging Face (2026): Building and deploying LLM-powered agents using open-source tools.

Education

B.Tech (Artificial Intelligence & Machine Learning): Sri Vasavi Inst. of Engg. & Tech., Nandamuru 2027
CGPA: 7.94

Intermediate: Pavithra Co-Op Junior College, Machilipatnam 2023
Percentage: 85.4%

SSC: Dr. KKR Gowtham School, Machilipatnam 2021
Percentage: 99%