

SUKESH S T

Chennai, India | stsukesh@gmail.com | +91-8870911769 | github.com/stsukesh | linkedin.com/in/sukeshst | leetcode.com/stsukesh

Career Objective

AI/ML Engineer specializing in Generative AI, Agentic AI, and full-stack AI applications. Experienced in building production ready intelligent systems using LLMs, LangGraph, RAG, and modern AI frameworks, with a strong focus on scalable, real-world solutions

Education

Bachelor of Engineering in Computer Science

Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai

CGPA: 8.15/10

2022 – 2026

Relevant Coursework: Data Structures, Algorithms, Machine Learning, DBMS, Networks, OOAD

Higher Secondary Education (HSC)

SRKBMHSS, Kulasekharam, Kanniyakumari

74.83%

2022

Internship

Machine Learning Intern, Flex Pvt Ltd(GitHub)

6 Jan – 20 Jan 2025

- Built Scratch and Dent Detection System using YOLOv11, Python, and OpenCV
- Reduced manual inspection time by 40% via AI-driven quality control
- Implemented real-time image processing for defect identification

Projects

Enterprise Research Agent (Live Demo | GitHub)

2026

- Engineered a stateful, multi-agent AI research pipeline utilizing LangGraph, coordinating Planner, Researcher, Extractor, Writer, and Reviewer agents to generate comprehensive reports
- Implemented a robust Human-in-the-Loop (HITL) checkpointing mechanism to pause execution, gather user feedback, and dynamically re-route the agent's research path
- Developed concurrent tool-calling capabilities (async search on Tavily, Arxiv, Wikipedia) and deployed on Hugging Face Spaces via FastAPI, SQLite, and Docker

Flex Policies RAG Chatbot (Live Demo | GitHub)

2026

- Built a Retrieval-Augmented Generation pipeline using LangChain, ChromaDB, and SentenceTransformers to query corporate policy documents with source attribution
- Developed a ChatGPT-style web interface with FastAPI backend, real-time similarity search, and markdown-rendered responses
- Deployed as a Dockerized application on Hugging Face Spaces with Groq-hosted LLaMA 3.1 for low-latency inference

Face Authentication System (1st Place, Internal SIH)

2024

- Designed secure facial authentication using Python, YOLO, and OpenCV
- Implemented advanced recognition algorithms for improved security

Technical Skills

Languages: Python, Java, C++, C

AI/ML: PyTorch, LangChain, LangGraph, RAG, SentenceTransformers, OpenCV, YOLOv11, NumPy, Pandas

Databases: MySQL, ChromaDB, FAISS

Web & APIs: FastAPI, HTML, CSS, JavaScript

Tools & Platforms: Docker, Hugging Face, Git, Uvicorn

CRM: Salesforce (Flows, Objects, Automation, Lightning Web Components)

Certifications

- Salesforce Platform App Builder (Simplilearn, Aug 2025)
- AWS Certification (Prepsinsta, 2025)
- Matlab Onramp (2024)

Professional Skills

- Leadership
- Communication
- Teamwork
- Problem-solving
- Adaptability