

VARUN MOTIYANI

📞 +91 9595661281 ✉ varun.motiyani14@gmail.com [🌐 LinkedIn](#) [🐙 GitHub](#)

About Me

Software Engineer with 3+ years of experience building scalable backend systems and high-performance APIs with Python, Django, and FastAPI. Skilled in applied GenAI/LLM engineering — RAG pipelines, agentic workflows, and evaluation frameworks — with a track record of accelerating product development and improving system efficiency.

Education

Vivekanand Education Society's Institute of Technology, Chembur, Mumbai	2020 - 2024
<i>B.E, Information Technology</i>	<i>CGPA : 9.57</i>

Experience

MicroStrategy Inc, IDC-Pune	July 2025 – Present
<i>AI Engineer 1</i>	

- **AutoAgents & AutoDash:**
 - * Sole architect of the **ReWOO + fallback ReACT** agentic architecture, a **tool-decoupled planning and execution pipeline** (web, geospatial, and proximity search modules) that solves complex interdependent queries, cutting response latency by **50%+** while improving accuracy and enabling single-point tool integration.
 - * Built end-to-end **Model Context Protocol (MCP)** integration with **fastmcp**, turning AutoAgents into a compliant **MCP client/host** and letting users connect to **Atlassian, GitHub, Notion, Mosaic**, and other MCP servers.
 - * Designed **SimpleMem**, a lightweight **agent memory architecture** (short/long-term, multi-tenant) backed by **ClickHouse** and **Redis**; shipped as a reusable **mstr-package** adopted across AutoAgents, AutoDash, and CustomerExpert.
 - * Sole owner of AutoDash's **ReAct-based orchestration engine** over parallel intent agents; fixed complex sequential-task failures by re-architecting it as a 2-stage async pipeline with fallback to the original approach.
- **AutoPrompt:**
 - * Built a standalone **microservice** for provider-aware prompt management with versioning and dynamic rendering by **model provider/variant**, making BYOLLM seamless across all AI products.
- **CLS AutoBot Smart Routing:**
 - * Built **ai-cls-router**, a **Python/FastAPI** case-routing monolith (**PostgreSQL, Redis, Azure OpenAI**) replacing a 6-service legacy system, auto-assigning engineers every 60s via a custom **FXScore workload algorithm** across **95+ REST APIs**, with real-time audit streaming over **Redis Streams/WebSockets**.
 - * Engineered a **custom support teams framework** (skill pools, overflow chains, safe-harbor restrictions) with **14 management endpoints**, and shipped production-grade security (**JWT RS256, RBAC, OIDC + PKCE**) integrated with **ServiceNow** and **Teams**, deployed on **AWS EKS**.
- **AI Driven EP Review Processing:**
 - * Automated the enterprise performance review cycle in **Python** with a **parallel processing architecture** handling 100+ reviews simultaneously, using **Azure OpenAI** to score feedback against competency frameworks and improving pipeline performance by **80%**.
 - * Built a multi-format ingestion pipeline (**video/Whisper, audio/MoviePy, documents**) with **Azure OpenAI**-based translation and evaluation, generating **DOCX** reports with automated **SMTP/Teams** bot notifications.

Saint-Gobain Indec, Mumbai	June 2024 – July 2025
<i>Associate Software Engineer</i>	

- **ServiceNow IT Support Chatbot:** Built an end-to-end **GenAI-powered chatbot** with agentic orchestration (**LangGraph/CrewAI**, open-source and **Azure OpenAI**) and a **Human-in-the-Loop** workflow, auto-resolving common IT issues and cutting resolution time by **40%**.
- **Confluence Knowledge Hub Chatbot:** Designed a **Hybrid RAG** system combining a **Neo4j** knowledge graph (LLM Graph Transformer, BFS/DFS, bi-entity matching) with traditional RAG, cutting employee query resolution time by **70%**.
- **Manufacturing Plants Analytics Chatbot:** Built an **LLM-based analytics chatbot** with **ML/time-series forecasting** for failure prediction over structured and unstructured plant data, deployed serverless on **AWS Lambda/S3** via **FastAPI**.

- Built a **Bi-LSTM based NER pipeline** with **OCR** and text preprocessing to extract key-value pairs from unstructured loan document PDFs, improving accuracy by 15% via **transfer learning**.
- Deployed the model using **Flask**, integrated with a **React** frontend, and used **PostgreSQL** for data storage on **Windows IIS**.

Projects

Fitness Tracker iOS

- Architected a provider-agnostic **LLMProvider** abstraction — a structured JSON request/response contract, never loose prose — with four swappable backend adapters (**OpenAI-compatible**, **Gemini**, **Anthropic**, on-device) selected at runtime via user-managed profiles, so new AI vendors or models plug in with **zero app updates**.
- Designed a **rule-engine + validator safety layer** that clamps every AI-generated training plan to hardcoded volume-landmark and progression rules, rejects invalid exercise IDs or unsafe load jumps, and **auto-retries then falls back** to deterministic rule-based planning on validation failure or API error.
- Building a real-time **per-call cost ledger** — token counts and cost computed from configurable provider pricing, with budget caps and usage dashboards — on a fully on-device data layer (**SwiftUI**, **SwiftData**); no backend, no accounts, single-user by design.

CareerConnect Portal

- Built an **ML-powered placement portal** using **Django** and **React**, featuring an **NLP-based ATS Resume Parser** (spaCy) to rank candidates against job descriptions.
- Developed a **Placement Prediction Module** using **Random Forest** and **XGBoost**, achieving 85% accuracy; deployed with **PostgreSQL** and **Docker**.

Skills

Languages: Python, Java, SQL, CypherQL, JavaScript

Machine Learning: Scikit-learn, TensorFlow, PyTorch, spaCy, NLTK, Transformers, Bert

Generative AI: Prompting, LLM (GPT, Claude, Open Source), LangChain, Agentic AI (LangGraph, CrewAI), RAG (Various Techniques), GraphRAG, Knowledge Graphs, LLM Testing and Evaluation (LangSmith, Opik, DeepEval), Simplemem

Data Science: Pandas, NumPy, Matplotlib, Seaborn, Time-Series Forecasting

Frameworks and Libraries: Flask, FastAPI, React, Node.js, Express.js

Database: PostgreSQL, MongoDB, Snowflake, Neo4j, ArangoDB, Redis, Vector DBs

DevOps: Docker, Kubernetes, Git, GitHub

Cloud: AWS (S3, Lambda, EC2), Azure (ADF, Synapse)

Certifications and Badges

Achievements

- **Anthropic Claude Certified Architect Foundational (CCAF)**.
 - **Oracle Cloud Infrastructure Gen AI Certified Professional**.
 - **Neo4j Certified Professional and Neo4j Certified Data Scientist**
 - **Snowflake Badges:** Data Warehousing, Data Engineering and Marketplace Badges.
- **Runner Up - Chairman’s Innovation Award’26**
 - **Q4 Best Engineer 2025 Microstrategy IDC AI**
 - **Q1 Best Engineer 2026 Microstrategy Global AI**