Devesh Sawant

AI Engineer | Full Stack Developer
Sangli, Maharashtra
9420707429 | deveshsawant300@gmail.com | deveshsawant12 | DeveshSawant12

Professional Summary

AI Engineer and Full Stack Developer with hands-on experience building Agentic AI systems, Retrieval-Augmented Generation (RAG) pipelines, Knowledge Graph applications, and scalable backend platforms. Experienced in developing autonomous AI agents, graph-memory architectures, and intelligent software products using Python, FastAPI, React, Node.js, Neo4j, Redis, vector databases, and modern LLM frameworks. Passionate about solving complex real-world problems through Generative AI, intelligent systems, and robust software engineering.

Education

Annasaheb Dange College of Engineering & Technology **2023 – 2027**
B.Tech in Artificial Intelligence & Data Science Engineering – CGPA: 8.73 *Ashta, Maharashtra*

Projects

AETERNA – Autonomous SRE Platform | *Python, FastAPI, Gemini, Groq, ChromaDB, MCP, React* **2026**

- Architected a tri-agent pipeline (Sentinel → Strategist → Validator) that autonomously detects, analyzes, and resolves infrastructure incidents end-to-end within 5 minutes with zero manual intervention.
- Developed RAG-based contextual memory (ChromaDB + Sentence Transformers) storing incidents as vector embeddings for history-aware decisions; engineered Redis alert deduplication via 60-second sliding window with sandbox-validated MCP action execution.
- Integrated predictive failure detection using historical incident patterns and trend analysis to proactively flag at-risk components before actual failures occur, reducing incident rate.
- Implemented a self-healing feedback loop that records action outcomes and progressively prioritizes higher-success remediation strategies, continuously improving resolution accuracy.

GraphMind – AI-Powered Interview Preparation Assistant | *Python, FastAPI, Neo4j, Milvus, Redis, LLMs, React* **2026**

- Built a personalized interview prep platform with a two-layer memory architecture (ephemeral TTL cache → Neo4j knowledge graph) that retains each user's learning history, weak areas, and progress across sessions.
- Designed a controlled memory promotion pipeline (Extract → Score → Metadata → Conditional Promote) to evolve user knowledge graphs over time without graph explosion.
- Implemented graph-first hybrid retrieval ($\leq$3-hop Neo4j traversal + Milvus vector search) for deterministic, evidence-backed answers to user queries with full explainability.
- Developed a VS Code extension tracking real-time coding behavior and an adaptive question engine that adjusts difficulty dynamically based on mastery score, improvement velocity, and contradiction-aware knowledge gaps.

Banking Automation and Verification Suite | *Python, OCR, ResNet, ELA, LangChain, Vector DB* **June 2025**

- Developed a RAG-based LLM chatbot for user queries and automated document verification; applied OCR and ResNet with Error Level Analysis (ELA) to detect forged Aadhaar/PAN documents with real-time fraud alerts via automated risk scoring.
- Enabled secure face verification and cross-document comparison for remote onboarding, ensuring end-to-end authenticity validation with backend fraud risk pipeline.

Technical Skills

Languages: Python, C++, C, HTML, CSS, Javascript, SQL
AI/ML Libraries: Scikit-learn, TensorFlow, Pandas, NumPy, ResNet, OpenCV, Tesseract OCR, LangChain, Sentence Transformers
Frameworks & Platforms: FastAPI, Flask, Neo4j, ChromaDB, Milvus, Redis, MySQL, Docker, Google Cloud Platform (GCP)
Developer Tools: VS Code, Git, GitHub, Google Colab, Jupyter Notebook, Postman
Expertise: Agentic AI, Generative AI, RAG & Graph Memory, LLM Engineering, Prompt Engineering, Machine Learning, Deep Learning, Computer Vision, OCR, NLP, Vector Databases, System Design

Achievements & Hackathons

Qualified – GATE 2026 (CS) | GATE Score: 366, AIR: 25,131, Marks: 31.33/100 **2026**
Winner – ADCET AI & DS Ideathon | AQI-Based Health Prediction and Alert System **2025**
Winner – RIT Hackathon | AI-Powered Green Energy Solution **2024**
Finalist – KIT College Hackathon | Green Energy Innovation Using Fruit Waste **2024**

Work Experience

Data Science Intern | Cognifyz Technologies **2026**

- Performed data preprocessing and exploratory data analysis (EDA) using Python and Pandas.
- Analyzed restaurant rating data, visualized distributions, and identified class imbalance.

Relevant Coursework

Machine Learning • Deep Learning • Web Development • Computer Vision • Operating System • Database Management Systems