SWAMI GAWAI

AI Engineer · Prompt Engineering · LLM Integration · Python · LangGraph · RAG
+91 9823275758 | swamigawai.official@gmail.com | Amravati, Maharashtra, India
linkedin.com/in/swami-gawai | github.com/swamigawai

PROFESSIONAL SUMMARY

AI/GenAI graduate with hands-on project experience building prompt-driven LLM systems, agentic workflows, explainable AI pipelines, and production-style AI applications across healthcare, customer experience, and payments domains. Built and deployed NexFlow AI Framework, a full-stack GenAI system using FastAPI, LangGraph, RAG, SQL, Docker, and AWS, and the Advanced AI Medical Intelligence Platform, a deep learning system for medical image classification with explainable AI and LLM-based clinical reporting. Strong foundation in prompt engineering, LLM behaviour, machine learning model development, REST APIs, data privacy and security practices, OOP, SQL, and technical documentation, with a focus on grounded outputs, model interpretability, and maintainable system design. Certified by LangChain Academy (Deep Agents, 2026), Anthropic Education (AI Fluency, 2026), and NVIDIA DLI (Generative AI Practice, 2025).

TECHNICAL SKILLS

Core Languages: Python, SQL, JavaScript
GenAI & LLMs: Prompt Engineering, LangGraph, LangChain, RAG, Hugging Face Transformers, LLaMA 3.2, Gemini
Backend & APIs: FastAPI, REST API Design, Async Python, Pydantic
Data & Storage: ChromaDB, FAISS, PostgreSQL, SQLite
Cloud & DevOps: AWS (Lambda, API Gateway, Bedrock), Docker, GitHub Actions, Git
ML & CV: PyTorch, CNNs (ResNet), Grad-CAM Explainability
Frontend: HTML5, CSS3, JavaScript, React (Fundamentals), Streamlit

PROJECTS

NexFlow AI Framework Mar 2026
AI / GenAI Project | github.com/swamigawai/NexFlow-AI-Framework

- Designed and iteratively refined system prompts, instruction structures, and few-shot examples to improve response quality across ambiguous and domain-specific user queries.
- Built a full-stack GenAI system covering data ingestion, document chunking, embedding, vector storage (ChromaDB), hybrid retrieval, re-ranking, and LLM response generation for grounded question-answering over private knowledge bases.
- Designed intent detection and query understanding logic in a LangGraph workflow to route user inputs to the correct retrieval or generation path and maintain grounded responses.
- Implemented an evaluation framework to test response quality across curated domain-specific queries and iteratively refined chunking strategy, retrieval settings, and prompt structure based on failure patterns.
- Deployed on AWS using Lambda, API Gateway, Bedrock, Docker, and GitHub Actions CI/CD with input validation and response safety checks; built the FastAPI async backend, the frontend, and a SQLite/PostgreSQL schema for conversation history and retrieval metadata.

Advanced AI Medical Intelligence Platform Jul 2026
AI / Deep Learning / Healthcare Project | github.com/swamigawai/Advanced-AI-Medical-Intelligence-Platform

- Built a full-stack medical AI platform that classifies pneumonia from chest X-rays using a PyTorch ResNet50 model fine-tuned via transfer learning, served through an asynchronous FastAPI backend.
- Implemented Grad-CAM-based explainable AI to generate visual heatmaps of model attention on radiographs, improving prediction interpretability and clinical trust.
- Integrated an LLM-based reporting pipeline (Gemini) with a custom fallback engine to auto-generate structured clinical reports (Findings, Interpretation, Recommendations) from model predictions.
- Engineered a HIPAA-aligned data security layer using SQLAlchemy ORM and Fernet AES encryption to protect Protected Health Information (PHI) at rest, applying best practices in data privacy for sensitive AI systems.
- Built a Streamlit clinical dashboard with an in-process fallback mode to maintain inference availability during backend outages, and containerized the full application with Docker for reproducible deployment.

Post-call Intelligence Orchestrator Jun 2026

Post-call Analysis Project | *github.com/swamigawai/post-call-intelligence-orchestrator*

- Built a post-call intelligence pipeline to extract call metadata, key intents, sentiment, and follow-up actions from recorded voice call transcripts using a mix of ASR, RAG, and LLM summarization.
- Designed extraction heuristics and prompt templates to surface caller issues, required follow-ups, and recommended agent responses; implemented evaluation checks to validate extraction quality.
- Integrated results into a lightweight dashboard and exposed REST APIs to retrieve summarized call insights for downstream CX tooling.

AI Payment Orchestration System May 2026
AI Automation Project | *github.com/swamigawai/ai-payment-orchestrator*

- Built a Python-based orchestration system to automate multi-step workflows, including request classification, routing, validation, retry logic, and status tracking.
- Applied OOP design principles, structured logging, and consistent error handling to make the workflow easier to trace, test, and maintain.
- Integrated REST APIs across workflow stages using typed data contracts and modular logic for easier extensibility.

EDUCATION

Bachelor of Engineering in Computer Science 2021 – 2025
HVPM College of Engineering, Amravati *CGPA: 8.16/10*

CERTIFICATIONS

- **Deep Agents: Agentic AI, LangGraph and LLM Orchestration** – LangChain Academy, Mar 2026
  academy.langchain.com/certificates/ecxly4uqw0
- **AI Fluency: Framework and Foundations** – Anthropic Education, 2026
  anthropic.skilljar.com/ai-fluency-framework-foundations
- **AI for All: From Basics to Generative AI Practice** – NVIDIA Deep Learning Institute, 2025
  verify.skilljar.com/c/n4dubahfazw6

LEADERSHIP

General Secretary, Utkarsh Forum 2024 – 2025

- Led a cross-functional team of 20+ members across 10+ concurrent initiatives, running sprint-style delivery cycles with zero missed deadlines while mentoring junior contributors and resolving blockers proactively.