ISHWARI WAKCHAURE

AI/ML Engineer | Pune, Maharashtra
+91-9561185411 | wakchaureishwari@gmail.com | LinkedIn

EXPERIENCE SUMMARY

AI Engineer with 2 plus years of experience building Generative AI solutions for clients in MedTech, Telecom, and Aerospace at Capgemini Engineering. Core expertise includes Python, LLM fine-tuning (LoRA, QLoRA, PEFT), RAG pipelines, multi-agent orchestration using LangChain and LangGraph, and cloud deployment on AWS. Proficient in Python, PySpark and deep learning frameworks including TensorFlow, PyTorch, and Keras, and API development using RESTful APIs with async/await patterns for production. Reduced operational timelines by up to 90+ percent through targeted AI automation. Recipient of two WOW Awards at Capgemini Engineering.

TECHNICAL SKILLS

Programming Languages: Python, C, SQL, Go

Frameworks and Libraries: TensorFlow, PyTorch, Keras, Scikit-Learn, LangChain, LangGraph, MCP, Azure OpenAI frameworks, Flask, OpenCV, Pandas, NumPy, Matplotlib, PySpark.

AI and ML: LLMs (GPT-4, Llama 3.1/3.2, Qwen Coder, Mistral), VLMs (Phi-3 Vision, Stable Diffusion), RAG, Prompt Engineering, Fine-Tuning (LoRA, QLoRA, PEFT), Agentic Systems, NLP, Deep Learning.

Cloud and DevOps: AWS, Azure, MCP, Docker, Kubernetes, CI/CD, MLflow

Tools and Platforms: FastAPI, REST APIs, Vector Databases (FAISS, ChromaDB, Neo4j), Databases (PostgreSQL, Redis), Git, Linux, Jupyter Notebook, JIRA

EMPLOYMENT PROFILE

Organization	From	To	Role	Specialization
Capgemini Engineering, Bangalore	Dec 2023	Present	Associate II Data and AI Engineer	AI / ML Generative AI

PROFILE SUMMARY

- Worked on multiple AI and ML projects across MedTech, Telecom, and Aerospace domains in Agile, collaborative environments, delivering analytical solutions to complex business problems.
- Specialized in Retrieval-Augmented Generation (RAG) pipelines, multi-agent orchestration using LangChain and LangGraph, agentic memory management, and AWS.
- Experienced in LLM fine-tuning using LoRA, QLoRA, and PEFT for domain-specific tasks. Worked with models including GPT-4, Llama 3.1/3.2, Qwen Coder, Phi-3 Vision, and Mistral.
- Built NLP solutions covering text classification, sentiment analysis, anomaly detection, and language generation, all optimized for production performance.
- Hands-on experience with Vision Language Models for multimodal tasks including UML diagram interpretation and visual document processing.
- Developed and deployed REST APIs (sync and async) using FastAPI for model serving and seamless system integration in production environments.
- Experienced with AWS and PySpark for big data processing, model training, and deployment using Docker and CI/CD pipelines.

- Strong communicator with experience preparing technical documentation, best-practice guides, and presenting findings to technical and non-technical stakeholders.
- Recognized for effective collaboration and delivering end-to-end AI solutions that consistently meet client requirements and deadlines.

PROJECT PROFILE

Project Name	Knowledge Graph Based Assistant for Medical Surgeon (Smith and Nephew)
Role	Developer

Responsibilities:

- Built a multi-agent system using LangGraph and LangChain with ReAct reasoning and adaptive memory, achieving 87 percent production accuracy on medical query responses.
- Designed an AWS Bedrock pipeline to convert unstructured multi-format documents into structured JSON, then indexed them using ChromaDB for Knowledge Graph-based retrieval.
- Added hallucination guardrails and factual grounding checks, reducing incorrect responses in medical documentation queries by over 60 percent.
- Used Pandas and NumPy for preprocessing and validating document data before ingestion into the vector store pipeline.

Project Name	Automated Unit Test Case Generation System (Ericsson)
Role	Developer

Responsibilities:

- Fine-tuned Qwen Coder using LoRA on telecom-specific Go, C, and Python codebases, improving automated test coverage by 60 percent.
- Built a FastAPI backend with structured prompt engineering pipelines, automated coverage reporting, and CI/CD integration so developers can generate and review test cases without manual effort.
- Used MLflow to track fine-tuning experiments, compare model checkpoints, and select the best-performing version before deployment.
- Wrote technical documentation and reusable prompt templates that were adopted by the broader engineering team. Received the WOW Award in 2024.

Project Name	Multi-Agent Surgical Log Analysis for Surgical Device (Smith and Nephew)
Role	Developer

Responsibilities:

- Built a dual-LLM pipeline using GPT-4 and Llama 3.1 that reduced surgical device log analysis time from 10 days to 20 minutes, a 90+ percent reduction.
- Connected JIRA tickets, system logs, and technical documents into a unified RAG pipeline with Human-in-the-Loop validation to support safe decision-making in medical workflows.
- Deployed the system on AWS using Docker, ensuring consistent performance across environments.
- Received the WOW Award in 2025 for delivering measurable impact through multi-agent orchestration and real-time AI diagnostics.

Project Name	Boeing Aerospace Chatbot

Role	Developer

Responsibilities:

- Built a document-grounded chatbot for a Boeing client using Llama 3.1 8B and Retrieval-Augmented Generation, enabling accurate responses from shared technical documents.
- Developed REST APIs using FastAPI and designed the user interface to allow smooth interaction between end users and the underlying LLM system.
- Delivered the complete solution on schedule, meeting all client requirements and integration expectations.

Project Name	Automated Python 2 to Python 3 Migration Tool (Internal Project)
Role	Developer

Responsibilities:

- Built an agentic pipeline using Llama 2 and LangChain that automatically migrates legacy Python 2 codebases to Python 3, handling syntax changes and project-specific conventions.
- Integrated automated unit test generation and coverage reporting using PyTorch-compatible test runners to confirm zero regression after each migration.
- Tracked all migration steps through Git version control, allowing rollback and diff comparison at any stage of the process.

Project Name	Test Case Generation from UML Diagrams (GYAN Platform)
Role	Developer

Responsibilities:

- Fine-tuned Phi-3 Vision, a vision language model, interpret UML sequence and state diagrams and generate structured test cases directly from visual design artifacts.
- Used OpenCV to preprocess diagram images before passing them to the model, improving recognition accuracy on complex UML layouts.
- Eliminated manual test case writing at the design phase, allowing QA teams to work directly from diagrams without needing to translate them into text-based descriptions first.

EDUCATION PROFILE

Degree / Course	Institution / University	Year	Result
B.E. - Computer Engineering	SPPU University, Pune	2023	CGPA: 9.23 / 10 (Distinction)

CERTIFICATIONS

IBM AI Engineer Professional Certificate - IBM

Agentic AI with LangChain and LangGraph - IBM

Deep Learning Specialization - DeepLearning.AI

AWARDS AND ACHIEVEMENTS

WOW Award (2025) - Awarded for building a multi-agent surgical log analysis system that cuts device investigation time from 10 days to 20 minutes. The system used GPT-4, Llama 3.1, and a RAG pipeline with Human-in-the-Loop validation to deliver real-time, safe diagnostics on medical devices.

WOW Award (2024) - Awarded for fine-tuning Qwen Coder using LoRA on telecom codebases, which raised automated test coverage by 60 percent and reduced manual testing effort for the client engineering team.