

NITHIN MOHAN NAIR

AI Developer

9792574524

Pune, India

linkedin.com/in/nithinnair01

nithin.ggplus1@gmail.com

ABOUT ME

Experienced AI Engineer at Tata Consultancy Services (TCS) with expertise in Generative AI, MLOps, and Responsible AI. Proficient in Python-based AI backends, fine-tuning Large Language Models (LLMs), and optimising inference pipelines. Skilled in deploying scalable AI solutions using cloud platforms (AWS, Azure) and MLOps/LLMOPs frameworks (MLFlow, NeMo Guardrails, AutoGen, LangChain). Passionate about building efficient, ethical AI-driven applications.

WORK EXPERIENCE

Tata Consultancy Services (TCS) — AI Engineer

2023–NOW

- Architected and scaled** Generative & Agentic AI backends in Python, leveraging Azure OpenAI GPT and Hugging Face Transformer models along with AutoGen and Google ADK Framework.
- Led development of an AI observability platform for LLMOPs and AgentOps**, generating actionable metrics and insights to optimize model performance, system reliability, and agent workflow efficiency.
- Designed and delivered** POCs/PoTs for **government and other group customers**, showcasing AI solution architectures—LLM-based chatbots, recommendation engines, and voice-enabled assistants—at industry events.
- Built Voice Agent interfaces** enabling natural-language interaction hence presenting a new way to interface with websites/solutions/data.
- Worked with **NVIDIA NeMo-Guardrails for Responsible AI** checks to ensure governance and compliance.
- Engineered high-performance inference pipelines with optimized vector search (FAISS/AWS Knowledge Bases by Bedrock) **cutting AI processing latency in existing solution by 40% (30–45 min → 7 min)** and boosting data throughput by 76.7%.
- Worked with established CI/CD pipeline** (GitLab, Jenkins, Kubernetes), gaining hands-on experience in continuous integration and deployment.
- Collaborated cross-functionally** with software engineering, DevOps, and domain stakeholders to integrate AI solutions into production platforms, driving an increase in operational efficiency while upholding data privacy and security best practices.

EDUCATION

JSPM'S RAJARSHI SHAHU COLLEGE OF ENGINEERING, PUNE

2019–2023

Bachelor of Technology (B.Tech) in Electronics and Telecommunication Engineering with **9.67 CGPA**.

GG INTERNATIONAL SCHOOL, PUNE

2018–2019

12th CBSE with **91% in PCM**.

GG INTERNATIONAL SCHOOL, PUNE

2016–2017

10th CBSE with **10 CGPA**.

SKILL

- **Programming & Development:** Python, Flask, FastAPI, Java, SQL, HTML/CSS
- **AI & ML:** Proficient in Generative AI and LLM workflows (fine-tuning, prompt engineering) using LangChain, Ollama, NeMo Guardrails, and MLflow. Hands-on building Agentic, voice-based, and multimodal agents with rapid prototyping and quick turnaround times. Worked with cloud AI platforms including AWS Bedrock, Azure OpenAI Services, GCP Vertex AI, and Gemini Services.
- **Cloud & DevOps:** Expertise in Azure, GCP and AWS and their services, containerization with Docker, and version control with Git/GitHub. Proficient in managing vector databases (FAISS) for scalable similarity search.

INTERNSHIPS & TRAINING

- **Intern at Tata Consultancy Services** Jan 2023 - May 2023
Data Engineering & ETL Framework in BFSI Unit
- **Microsoft Azure Virtual Internship** Nov 2021 - Feb 2022
Built a Campus Guide Chatbot using Azure QnA Maker
- **Zensar ESD Training** Sept 2021
Professional Communication, SQL-PLSQL, and Testing
- **SP&IR Cell Internship Digital Intern** Aug 2021 - Feb 2022
Digital Intern for social media content & video editing

VALUE PROPOSITION

- **Quick RFP Analysis:** Break down customer requirements and key pain points into a clear technical plan within 48 hours of receiving an RFP.
- **End-to-End Design:** Draft scalable AI/ML solution outlines—including data pipelines, model selection, deployment approach, and monitoring strategy—with minimal ramp-up time.
- **Fast Prototyping:** Develop functional POCs or PoTs in under a week using existing frameworks (AutoGen, LangChain, NeMo) and cloud AI services to demonstrate feasibility.
- **Clear Stakeholder Communication:** Create concise solution briefs and architecture diagrams that link technical choices directly to business goals, ensuring alignment across technical and non-technical teams.

REWARDS AND RECOGNITIONS

- **Star of the Month |** Improvements in Platform and Creation of Platform from Scratch – TCS, Dec 2024
- **Technical Excellence Award |** Agentic AI – TCS, June 2025