

Sanjeev Kumar

Noida, India | +91 7289909687 | sksanjeevkumar489@gmail.com | linkedin.com/in/sanjeev-k-6b2490b9

OBJECTIVE

Generative AI Engineer with 6.5 years of experience designing, developing, and deploying scalable AI solutions on cloud-native platforms including Azure, AWS, and Google Cloud. Proven expertise in LLM-powered applications, prompt engineering, Agentic AI, and cloud infrastructure automation. Seeking to deliver innovative GenAI solutions blending cutting-edge AI research with production-grade cloud scalability.

SKILLS

- Generative AI & ML: Python, Prompt Engineering, LLMs (GPT-3.5, GPT-4, LLaMA), LangChain, Agentic AI, LangGraph, RAG, OpenAI API, Microsoft Copilot Studio
- Cloud Platforms: Azure (OpenAI, Functions, Cognitive Services, Copilot Studio), AWS, Google Cloud (Vertex AI)
- Frameworks & Tools: Terraform, Docker, crewAI, AI Refinery, FastAPI
- Databases: CosmosDB, PostgreSQL, MongoDB, Pinecone Vector DB, Redis

PROFESSIONAL EXPERIENCE

Generative AI Engineer

Tata Consultancy Services (TCS) — Jan 2025 – Present | Noida, India

- Designed and deployed enterprise GenAI chatbots and workflow automation solutions leveraging **Python**, **OpenAI GPT-4**, **LangChain**, and **Microsoft Copilot Studio**.
- Implemented agentic AI automation workflows, integrating Microsoft Copilot Studio for rapid agent development and orchestration.
- Architected cloud-native AI deployments on Azure and Google Cloud (**Vertex AI**) to maximize scalability, fault tolerance, and performance.
- Applied advanced prompt engineering methodologies (**zero/one/few-shot, RAG**) to improve user experience and system accuracy.
- Automated AI solution delivery pipelines using Terraform, Docker, and serverless Azure Functions.

Generative AI Engineer

Capgemini — Jan 2024 – Jan 2025

- Developed and maintained knowledge **agents** and chatbots using **OpenAI**, **LangChain**, and **Google Vertex AI**, improving customer interaction automation.

- Automated cloud infrastructure provisioning using Terraform for scalable GenAI services on Azure and Google Cloud.
- Led AI pipeline CI/CD incorporating model training, testing, and deployment.
- Pioneered adoption of **Microsoft Copilot Studio** for enterprise agent prototypes and AI service integration.

Generative AI Engineer

Infosys — Dec 2021 – Jan 2024

- Built production-grade GenAI assistants utilizing **LangGraph** and **Azure cloud AI services**.
- Integrated various cloud databases and vector search engines for real-time **LLM retrieval-augmented generation (RAG)**.
- Awarded Infosys Insta Award for DevOps and Generative AI automation excellence.

Junior DevOps Engineer

NTT DATA — Jul 2019 – Nov 2021

- Automated cloud provisioning and container orchestration pipelines, supporting AI data workloads.

GENAI PROJECT HIGHLIGHTS

- Delivered highly scalable, production-level GenAI chatbots with GPT and LangChain, optimized through advanced prompt engineering.
- Led automation and integration of large-scale AI workflows using Microsoft **Copilot Studio** and **Google Vertex AI**.
- Implemented seamless AI application deployment using Terraform on multi-cloud environments.
- Developed robust vector search systems integrating Pinecone, CosmosDB, and relational databases for AI applications.

CERTIFICATIONS & BADGES

- **Google Certified Generative AI Leader**
- **Microsoft Skilled Badge in Agentic AI**
- AWS Certified Cloud Practitioner & Microsoft Certified: Azure Fundamentals (AZ-900)

EDUCATION

B.Tech, Electronics & Communication Engineering

Dr. A.P.J. Abdul Kalam Technical University (2015–2019)

