

# Sahana Khanai

## Generative AI Engineer

[LinkedIn](#) | 9353048023 | [sahana.khanai.jcer.cs061@gmail.com](mailto:sahana.khanai.jcer.cs061@gmail.com) | Mumbai, India 400601 | [GitHub](#)

---

### Summary:

Generative AI Engineer with 2 years of experience developing enterprise-ready AI applications using LLMs, LangChain, Hugging Face, and RAG pipelines. Skilled in integrating vector databases (FAISS, Pinecone) and deploying Transformer-based solutions on AWS (SageMaker, Bedrock, Lambda). Specialized in multi-agent systems, prompt engineering, and fine-tuning models for domain adaptation. Strong background in MLOps, CI/CD, Agile methodologies, and scalable cloud deployment, with a proven record of reducing hallucinations, improving automation efficiency, and delivering production-grade AI solutions that accelerate business outcomes.

### Education:

- Bachelor of Engineering | Computer Science and Engineering** August 2019 – June 2023  
Visvesvaraya Technological University – Belagavi, **CGPA: 9.14**
  - Recipient of **Gold Medal** 2023

### Skills:

- Generative AI & NLP:** LLMs (GPT-4, Llama2, Mistral, Claude, Gemini Pro), LangChain, LangGraph, Hugging Face, RAG Pipelines (Retrieval-Augmented Generation), Prompt Engineering, Instruction Tuning, Fine-tuning, LangSmith, LangServe, LlamaIndex, Ollama, Text Embedding, Conversational AI, NLP Pipelines, NLTK, word2vec, Foundation Models, NVIDIA NIM, CrewAI
  - Machine Learning & Deep Learning:** NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, PyTorch, Keras, Scikit-learn, ANN, RNN, LSTM, Transformers, Transfer Learning, Model Optimization, Model Evaluation, Deep Learning Pipelines, OpenAI API
  - Cloud Platforms & MLOps:** AWS SageMaker, AWS Bedrock, Databricks, Docker, Jenkins, Git, SonarQube, CI/CD Pipelines, Unix, Model Deployment, LLMOps, MLflow, Agile Methodologies.
  - Full-Stack & API Development:** Spring Boot, RESTful APIs, FastAPI, Streamlit, OpenAPI (Swagger), Postman, Backend-Frontend Integration
  - Frontend Development:** React.js, JSP, JavaScript, HTML, CSS, Bootstrap
  - Databases & Vector Stores:** PostgreSQL, MySQL, Apache Cassandra, Astra DB, FAISS, Pinecone, ChromaDB, Query Optimization, Vector Embeddings
  - Programming Languages:** Python, Java, JavaScript, SQL
  - Soft Skills** - Ability to break down complex problems into structured solutions, clearly present ideas across technical and non-technical teams, collaborate effectively in group settings, and quickly adapt to new tools and technologies.
- 

### Experience:

#### Tata Consultancy Services (TCS)

##### Software Engineer – Generative AI Engineer

Mumbai, India | Feb 2024 – Present

- Engineered GenAI solutions with GPT-4, Llama2, and Mistral using LangChain & Hugging Face, boosting accuracy by 28%.
- Architected RAG applications with FAISS & ChromaDB, cutting hallucinations by 30% and latency by 25%.
- Orchestrated multi-agent workflows in LangGraph, improving automation efficiency by 35%.
- Deployed cloud-native AI systems on AWS (Lambda, Bedrock), automating processes and reducing manual effort by 40%.
- Fine-tuned LLMs on custom datasets, achieving 20% improvement in domain adaptation.

## **NSREEM**

### **AI Developer – Full Stack & ML Prototyping**

*Remote | Feb 2023 – May 2023*

- Created interactive Streamlit dashboards with Python, increasing engagement by 20%.
- Deployed clustering models for customer segmentation, improving marketing efficiency by 18%.
- Executed sentiment analysis with NLTK & Scikit-learn, improving classification accuracy by 22%.
- Prototyped a recommendation engine with OpenAI APIs & vector retrieval, achieving 15% accuracy gain.

## **Knowx Innovations Pvt. Ltd.**

### **AI/ML Engineer – Applied Machine Learning Projects**

*Bengaluru, India | Aug 2022 – Sep 2022*

- Delivered fraud detection with Random Forest, achieving 95% accuracy and reducing false positives by 20%.
  - Streamlined text classification workflows, cutting manual effort by 30%.
- 

## **Projects:**

### **1. ThreatLens – GenAI-Powered Cyber-Threat Assistant**

- Engineered a multi-tab LLM assistant to streamline cybersecurity workflows, cutting analyst triage time by 35%.
- Automated phishing detection with IOC extraction and remediation steps, improving detection efficiency by 25%.
- Created a CVE explainer translating vulnerabilities into plain English, increasing accessibility for non-technical users by 40%.
- Enabled analysts to process logs 50% faster via contextual summarization with actionable insights.

### **2. SOP-Genius – RAG-Based SOP Knowledge Assistant**

- Architected a Q&A assistant that converted static SOPs into a searchable knowledge base, improving query resolution time by 45%.
- Integrated FAISS-based vector retrieval for context-aware responses, reducing irrelevant outputs by 30%.
- Launched a Streamlit interface that simplified navigation and accelerated adoption across cybersecurity teams.

### **3. Multi-Agent Logistics Chain Optimizer**

- Spearheaded a multi-agent logistics optimizer delivering 30% higher delivery efficiency through real-time decisioning.
  - Orchestrated specialized agents for orders, inventory checks, route optimization, and issue resolution.
  - Integrated RAG with FAISS for precise retrieval, cutting routing errors by 25%.
  - Deployed an interactive Streamlit dashboard for route visualization and explainable agent reasoning.
- 

## **Certifications:**

- Generative AI Foundation Curriculum – TCS
  - Complete Generative AI Course with LangChain & Hugging Face – Udemy
  - Career Essentials in Generative AI – Microsoft & LinkedIn
  - Introduction to OpenAI & ChatGPT API – Udemy
  - Machine Learning for All – University of London (Coursera)
  - SQL for Data Science – University of California (Coursera)
  - Python Data Structures – University of Michigan (Coursera)
- 

## **Achievements:**

- Gold Medalist – B.E. Computer Science, VTU (2023)
- Special Initiative Award – TCS (2025)
- Beyond Performance Award – TCS (2024)
- 5-Star Python Programmer – HackerRank