# Sahana Khanai

**Generative AI Engineer**

LinkedIn | 9353048023 | sahana.khanai.jcer.cs061@gmail.com | Mumbai, India 400601 | GitHub

## Objective:

Innovative and technically versatile Generative AI Engineer with 2 years of experience in designing intelligent applications using LLMs, LangChain, Hugging Face, and RAG pipelines. Skilled in integrating vector databases such as FAISS and Pinecone, and deploying Transformer-based models on cloud platforms including AWS SageMaker, Bedrock, and Lambda. Combines expertise in multi-agent systems, prompt engineering, and fine-tuning with strong knowledge of MLOps, model deployment, and modern AI frameworks to deliver scalable, production-ready AI/ML solutions in enterprise environments.

## Education:

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

## Skills:

- **Generative AI & NLP -** Large Language Models (LLMs) – OpenAI GPT, Llama2, Mistral, Gemini Pro; LangChain, Hugging Face, Transformers, RAG Pipelines (Retrieval-Augmented Generation), PEFT, Prompt Engineering, Instruction Tuning, Fine-tuning with custom datasets, LangSmith, LangServe, LlamaIndex, Ollama, NLP Pipelines, NLTK, word2vec, Text Embedding, Conversational AI, Foundation Models, NVIDIA NIM, CREW AI
- **Machine Learning & Deep Learning -** TensorFlow, PyTorch, Keras, Scikit-learn, ANN, RNN, LSTM, Deep Learning Pipelines, Model Inference, Model Evaluation, Neural Network Optimization, Transfer Learning, OpenAI API
- **Cloud Platforms, MLOps & DevOps -** Amazon SageMaker, AWS Bedrock, Databricks, Docker, Jenkins, Git, CI/CD Pipelines, SonarQube, Unix, Git.
- **Full-Stack Development & API Engineering -** Spring Boot, Spring Framework, Hibernate ORM, RESTful APIs, Servlets, JSP, JDBC, Streamlit, OpenAPI (Swagger), Postman, Backend-Frontend Integration
- **Frontend -** React.js, JavaScript, HTML, CSS, Bootstrap
- **Databases** - PostgreSQL, MySQL, Apache Cassandra, Astra DB, Chroma DB, FAISS, Query Optimization, Vector Stores, Pinecone.
- **Programming** - Python, Java, JavaScript, SQL
- **Soft Skills** - Analytical Thinking, Creative Problem-Solving, Effective Communication, Team Collaboration, Adaptability, Fast Learning, Initiative, Time Management, Curiosity, Attention to Detail

## Experience:

**Tata Consultancy Services (TCS)**

**Software Engineer – Full Stack Development & Generative AI** *Mumbai, India | Feb 2024 – Present*

- Designed and implemented Generative AI solutions using LLMs (GPT-4, Llama2, Mistral, Claude) with LangChain, LangGraph, and Hugging Face Transformers.
- Built RAG-based applications in Python, integrating vector databases (FAISS, ChromaDB) to enable context-aware document retrieval and reduce hallucinations.
- Developed and tested multi-agent workflows with LangGraph for task orchestration, decision-making, and explainable AI pipelines.
- Delivered cloud-native AI solutions leveraging AWS Lambda, S3, and Bedrock, including automated content generation workflows.
- Applied prompt engineering, instruction tuning, and fine-tuning with custom datasets to improve response quality and domain adaptation of LLMs.

- Deployed prototypes and POCs with Streamlit, REST APIs, and Docker, ensuring scalability and enterprise integration.

**NSREEM**
**AI Developer – Full Stack & ML Prototyping** *Remote | Feb 2023 – May 2023*

- Created interactive, data-driven dashboards using Streamlit and Python, improving user engagement and system insights.
- Developed K-Means clustering models for customer segmentation and integrated them into Flask-based APIs
- Implemented basic sentiment analysis for customer feedback, using NLTK and Scikit-learn.
- Prototyped a content-based recommendation engine and explored GenAI integrations using OpenAI APIs and vector similarity.

**Knowx Innovations Pvt. Ltd.**
**AI/ML Engineer – Applied Machine Learning Projects** *Bengaluru, India | Aug 2022 – Sep 2022*

- Built a secure Password Manager application using Python, Tkinter, and MySQL, applying encryption and GUI principles.
- Developed a fraud detection system using a Random Forest classifier, achieving 95% accuracy on financial datasets.
- Preprocessed datasets, conducted feature selection, and trained models using Scikit-learn, improving model performance metrics.
- Explored NLP techniques and began implementing text classification workflows for internal automation use cases.

---

## Projects:

**1. ThreatLens – GenAI-Powered Cyber-Threat Assistant**
*Streamlit | OpenRouter API | Mistral | Python*
Built a multi-tab web assistant that harnesses Large Language Models to streamline security analysts' workflows.

- Phishing Analyst: flags suspicious emails, extracts IOCs, and suggests remediation steps.
- CVE Explainer: translates CVE details into plain English, outlining impact and mitigation.
- Log Summarizer: condenses lengthy security logs into actionable insights for faster triage.
- Integrates OpenRouter LLM endpoints within an intuitive Streamlit UI to deliver contextual, on-demand threat intelligence.
- Accelerates incident response and reduces analyst workload by providing cohesive AI support in one place.

**2. SOP-Genius – RAG-Based SOP Knowledge Assistant**
*LLMs | FAISS | RAG | Claude| Python*
Developed a smart Q&A assistant that transforms static cybersecurity SOPs into an interactive, searchable knowledge system using Retrieval-Augmented Generation (RAG).

- Enables users to upload or query SOP documents and get precise, context-aware answers powered by Large Language Models.
- Utilizes FAISS for efficient vector search and document retrieval.
- Built with a clean Streamlit interface for seamless user interaction.
- Helps cybersecurity teams quickly access relevant procedures, enhancing operational efficiency and decision-making.

**3. Customer Churn Classifier**
*Streamlit | ANN | TensorFlow | Keras | Python*
Developed an interactive web application to predict customer churn using an Artificial Neural Network (ANN).
Enabled businesses to proactively retain customers by accurately identifying churn risk, with a clean UI built using Streamlit.

- Achieved >85% prediction accuracy on test data.
- Enabled proactive customer retention strategies.
- Showcased end-to-end ML lifecycle: preprocessing, training, evaluation, and deployment.

**4. Movie Review Sentiment Detector**

*Streamlit | NLP | RNN | Python | TensorFlow*

Built a sentiment analysis tool using a Simple RNN to classify movie reviews as positive or negative. Demonstrated the use of deep learning for natural language understanding with interactive web UI.

- Achieved 90% sentiment classification accuracy.
- Visualized predictions and model confidence in real-time.
- Applied tokenization, embedding layers, and sequential modeling.

**5. Next Word Predictor**

*Streamlit | LSTM | NLP | TensorFlow | Keras*

Engineered an intelligent word prediction tool using LSTM networks to suggest the next probable word based on previous input, simulating natural typing behavior.

- Demonstrated sequence modeling and text generation capabilities.
- Improved language understanding using pretrained embeddings.
- Built a performant backend optimized for sequential data.

**6. Multi-Agent Logistics Chain Optimizer**

*LangGraph | LangChain | Mistral | RAG | Python | Streamlit | FAISS*

Developed a **multi-agent system** to optimize logistics operations, leveraging LangGraph and LLMs for real-time decision-making and explainability.

- Designed specialized agents for order management, inventory checks, route optimization, delay handling, and issue resolution.
- Integrated RAG-based retrieval with FAISS to provide agents with accurate, context-aware information.
- Improved delivery success rate and efficiency through automated rerouting, fallback handling, and dynamic scheduling.
- Built an interactive Streamlit dashboard to visualize routes, agent interactions, and reasoning steps for transparency.

---

## Certifications:

- Complete Generative AI Course with Langchain and Huggingface -Udemy
- Career Essentials in Generative AI – Microsoft & LinkedIn
- Introduction to OpenAI & ChatGPT API – Udemy
- Generative AI Foundation Curriculum – TCS
- SQL for Data Science – University of California, Coursera
- Machine Learning for All – University of London, Coursera
- Python Data Structures – University of Michigan, Coursera
- Cybersecurity Awareness & Foundations – TCS & Forage

---

## Achievements & Awards:

- Achieved **Gold Medal in the B.E. Computer Science department**.
- Recognized as the **Best Student** in high school and PUC.
- Secured **1st rank** in Computer Competition and SSLC exam, earning accolades from the high school.
- **5 Star** in Python programming in HackerRank, **Top performer** in **ILP TCS training program**.
- In TCS received **Beyond Performance Awards, Xcelerate Warrior Certificate.**
- **Special Initiative Award** from **TCS** | June 2025
- Recognized for outstanding contributions and proactive initiatives that positively impacted the team and organization. Appreciated as a role model for dedication and commitment.