

ROHAN SITANIYA

Senior Agentic AI Engineer | rsitaniya97@gmail.com | +91 7407997776 | [LinkedIn](#) | [Website](#)

6 years of building production grade intelligent systems across SaaS, finance, retail, and healthcare, built for the places where trust is non-negotiable: Fraud And Credit Risk, Underwriting and Governance at American Express, Agentic Workflow Automation at Jitterbit, Genomic Cancer Detection back at IIT Kharagpur. Give me noisy data and a vague business problem, I'll find the signal and deliver the system that acts on it.

TECHNICAL SKILLS

LLM & GenAI: Agentic AI, RAG, Multi-Agent Orchestration (LangGraph, CrewAI, AutoGen), Function Calling, MCP, Context Engineering, OpenAPI Generation, Structured Outputs (Pydantic), LLMops, AI Guardrails, LLM-as-Judge, DeepEval, Langfuse, AI Observability, Async Streaming, Semantic Retrieval, Vector DBs (Pinecone, Qdrant), Ollama, HuggingFace, OpenAI, Bedrock

Backend & Distributed Systems: FastAPI, Microservices, Kubernetes, Docker, Celery, Redis, Kafka, Apache Spark, Async Workflows, CI/CD Pipelines, REST APIs, Workflow Orchestration

Cloud & Data: AWS, GCP, PostgreSQL, MongoDB, MySQL, Python, OpenCV, Spark, C++

PROFESSIONAL EXPERIENCE

Senior Agentic AI Engineer - Jitterbit

Oct 2024 – Present

Enabling Citizen Developers through Enterprise Workflow automation using Agentic AI.

Remote

- **iPaaS Bot:** Architected and built a multi-agent system for **Enterprise Workflow Automation** leveraging LangGraph, agent toolchains, function calling, and resumable checkpointed execution. Integrated RAG in the planning agent for improving context precision by over 60% and reduced retrieval latency by 30%; This enables all business and non-tech teams to build, manage and automate their workflows on Jitterbit's Platform with ease and confidence using just natural language.
- **APIM Bot:** Built a natural language driven **API lifecycle management Agent** serving 10K+ daily interactions; enabling API creation, configuration, publishing, testing and deletion.
- **AI, Responsible AI & Platform Engineering:** Implemented governance through DeepEval guardrails, hallucination/bias detection, Langfuse tracing, model drifting, reducing latency, zero-sensitive-data exposure patterns, and CI/CD-backed Kubernetes microservices for regulated enterprise environments. Context Engineering and conversational memory summarization, Pydantic-based structured outputs, async streaming pipelines, and provider-agnostic LLM orchestration (OpenAI, Bedrock, OSS models).

AI Engineer - American Express (AI Labs; Emerging AI Products; Model Risk Management)

Aug 2020 – Sep 2024

Built NLP and GenAI solutions for risk, fraud, underwriting, and GenAI governance

Full-time

- **GenAI Governance:** Led a 5-member team to build an enterprise GenAI governance framework with 20+ quantitative and 30+ qualitative evaluations covering hallucination, bias, toxicity, PII, faithfulness, auditability, observability. Integrated open source (DeepEval, Presidio, RAGAs, Ares) and custom built code for automated model gating, **reducing model governance time by ~60%**.
- **Risk Intelligence Engine:** Built an LLM-powered financial risk monitoring pipeline leveraging real-time news and external signals to automate early warning alerts for credit risk teams, enabling them to reduce loss margins.
- **AI Underwriting Automation:** Developed and deployed NLP based transaction categorization models (~85% accuracy) for bank-statement underwriting, **significantly reducing manual review cycles** of new account applications from days to hours.
- **Complaint Intelligence Platform:** Built a unified analytics platform for Complaints Resolution; combining product reviews on social media, call transcripts, and internal customer data to identify and help resolve consumer pain points and business trends.
- **Fraud Embeddings:** Designed Transformer-based embedding architectures outperforming legacy RNN models by ~40bps and developed GPT-based embeddings for large-scale fraud detection pipelines.

ML Engineer Intern - Algonomy

May 2019 – Jul 2019

- **Retail AI:** Fine-tuned BERT & GPT-2 for personalized query auto-completion on corpus of 1.5 million business analytics queries.

Research Intern — National Taiwan University, Epigenetics Lab

May 2018 – Jul 2018

- **Genomic analytics** of Axolotl RNA-Seq datasets to identify regenerative epigenomic markers and their expression pattern and map to human genome for enhanced regeneration.

EDUCATION

Integrated Dual Degree (B.Tech + M.Tech - Biotechnology) - IIT Kharagpur CGPA: 7.64/10

2015 – 2020

Thesis: Non-invasive detection of human cancers with multi-genomic TCGA data using Deep Learning.