

ANIRUDDHA SHRIKHANDE

shrikhandeaniruddha@gmail.com [Linkedin](#) 9890071283

AI Systems Engineer specializing in scalable Agentic Architectures, low-latency RAG systems, and production LLM infrastructure. Designed and deployed distributed AI systems serving 200K+ users with a focus on performance optimization, evaluation reliability, and cost-efficient inference.

SKILLS

AI Systems and Architecture

Multi-Agent Systems, RAG, LLM Evaluation, Guardrails, Fine-tuning Prompt Optimization, LLM observability, Red Teaming.

LLM Infrastructure and Deployment Data and Retrieval Systems

LangGraph, LangChain, LlamaIndex, MCP, ADK, Docker, Kubernetes. ChromaDB, Pinecone, Neo4j, SQL, Hybrid Retrieval, Embedding Optimization

Cloud and Backend

Python, FastAPI, Flask, GCP, AWS, PostgreSQL

EXPERIENCE

Associate AI Specialist — Association of Data Scientists (ADaSci)

October 2024 – Present

Client Projects

Advanced Micro Devices [AMD] (*Developer Engagement Platform*)

- Architected a **gamification engine** with **13 activity types**, **tiered referral rewards**, and **multi-layer deduplication** — powering real-time leaderboards.
- Designed an **admin analytics dashboard** with **precomputed PostgreSQL views**, a **7-stage conversion funnel**, enabling sub-second insights across 100k users.
- Engineered a **serverless social media verification pipeline** using **Edge Functions** to auto-verify posts across LinkedIn, X, and YouTube — eliminating 90% of manual review.

Tata Elxsi (*AI-Based Coding Evaluation & Technical Screening Platform*)

- Engineered a **custom boilerplate generation engine** supporting 5+ languages that auto-scaffolds ready-to-code environments based on question context, **saving candidates ~15 minutes** of setup per session.
- Designed a seniority-banded AI evaluation framework with deterministic scoring rubrics and validation checks to minimize grading variance and hallucinated feedback.

EXL (*Gamified Learning Platform*)

- Designed and automated the generation of **high-fidelity synthetic datasets** across finance, healthcare, and retail, simulating real-world business scenarios for AI-driven case studies and model evaluation.
- Built scalable **data generation pipelines** with domain-specific rules and validation layers, reducing manual dataset preparation time by ~70% while ensuring schema consistency and statistical realism.

Internal Projects

- Architected a real-time AI video interview system with streaming avatar responses, structured LLM scoring pipelines, async FastAPI backend, and resilient S3 multipart uploads with crash recovery — ensuring consistent session handling under unstable network conditions.
- Authored 70+ high-ranking tutorials and co-developed the Certified Agentic AI Certification (50K+ learners impacted), also delivered hands-on workshops on ADK and MCP, driving large-scale developer adoption.

PROJECTS

Multimodal Clinical AI Assistant([Github Link](#))

- Built a multimodal AI assistant using **DeepSeek** and **Janus** for text, PDF, image, and audio inputs
- Architected a low-latency RAG pipeline with hybrid retrieval and evaluation guardrails, improving answer accuracy by 40% while maintaining sub-700ms response time in production.
- Fine-tuned models with LoRA on med Q&A and VQA, improving accuracy by 85%
- Enabled features like **lab report comparison**, **multi-doc ingestion**, **image analysis**, and **audio Q&A**.

PUBLICATION

Color and Texture Based Similarity Search to Enable Secure Image Sharing on Social Media Platforms.

2023 3rd Asian Conference on Innovation in Technology.(ASIANCON)(**IEEE**)([Link](#))

EDUCATION

B.Tech in Electronics and Telecommunications, Vishwakarma Institute of Technology, Pune

2020 - 2024

CGPA:-8.34/10.0