

KUNAL SHARMA

+91 9648031505 ◇ Faridabad, Haryana, India

roshians@proton.me ◇ linkedin.com/in/roshians ◇ github.com/roshians

SUMMARY

Undergraduate AI/ML student with hands-on experience building multi-agent LLM systems, RAG pipelines, and production AI workflows using Python, FastAPI, and cloud services. Experienced with LLM APIs (Gemini, AWS Bedrock), vector search, and async programming. Proven through research internships at Wildlife Institute of India, Kaggle competitions, and production-grade AI projects serving real users.

EDUCATION

B.Tech in Computer Science (AI & ML)

2023 – 2027

IILM University, Greater Noida

SKILLS

Languages	Python (async/await), SQL, C
AI/ML & Agents	LLM APIs (Gemini, OpenAI, Anthropic), Multi-Agent Systems, RAG, Prompt Engineering
Vector & Search	FAISS, ChromaDB, Pinecone (basic), Embedding Pipelines, Semantic Search
Data & Vision	PyTorch, TensorFlow, Scikit-learn, OpenCV, CNNs, NLP, Pandas, NumPy
Cloud & Tools	AWS (Bedrock, Lambda), FastAPI, Streamlit, PostgreSQL, Git

EXPERIENCE

AI/ML Intern

Mar 2026 – Apr 2026

[Wildlife Institute of India \(WII\)](#) — *Elephant Re-Identification (Computer Vision)*

- Built an open-set visual re-identification system using ConvNeXt-Tiny with part-based pooling, generating 384-D embeddings for 259+ identities — directly applicable to image-based skin analysis and diagnosis workflows.
- Achieved Rank-1 accuracy of 92.16% and Rank-5 accuracy of 100% using Test-Time Augmentation (TTA) on unseen identities; false merge rate of 0.42% at 0.75 threshold.
- Performed end-to-end data preprocessing and augmentation on large unstructured image datasets; conducted EDA to guide model design decisions.
- Enhanced feature discrimination using Triplet, ArcFace, and Center Loss for robust identity separation.

PROJECTS

PathShala AI – Multi-Agent AI Co-Teacher

[GitHub](#)

- Engineered a multi-agent LLM system using Gemini API with a RAG retrieval pipeline and FAISS-based vector search; designed agent handoff logic to route queries between specialist agents based on context and intent.
- Built async FastAPI endpoints for multi-channel delivery (Android, WhatsApp, telephony) integrated with AWS Lambda, enabling scalable real-time AI interactions for classrooms with 5 grades and 60+ students.
- Applied NLP techniques for text preprocessing and content extraction; structured prompts for accuracy, relevance, and coherence – delivering context-aware responses in under 15 seconds.

BimaBot – AI-Powered Insurance Claim Auditor

HackVeda finalist (IBM & AWS)

- Built a document-intelligence platform where users upload hospital bills and insurance policy PDFs; used AWS Textract to extract structured text from complex 50-page policy documents.
- Designed a multi-document RAG pipeline on AWS S3 integrating three knowledge sources — user bill, insurance policy, and IRDAI government guidelines — enabling context-aware cross-document reasoning via AWS Bedrock.
- Generated structured overcharge reports flagging bill line items that violate policy terms or IRDAI caps, and produced ready-to-send dispute letters — advancing to Round 2 finals.

ACHIEVEMENTS

- Achieved top 25% ranking in Kaggle competition – applied structured data analysis, feature engineering, pattern recognition, and iterative model refinement.

CERTIFICATIONS

- [Introduction to Agentic AI on AWS](#) – AWS Jan 2026
- [Introduction to Generative AI](#) – AWS Educate Jan 2026
- [Artificial Intelligence Fundamentals](#) – IBM Jul 2025