KUNAL KUMAR SHAW

Raichur, Karnataka, India

📞 +91-6290308120 | ✉ kunalshawkol17@gmail.com | 🔗 LinkedIn | ⌾ GitHub | 🌐 Portfolio

PROFESSIONAL SUMMARY

Computer Science undergraduate driven by the engineering challenges behind Generative AI—specifically making LLMs efficient, reliable, and deployable. My focus lies in optimizing RAG pipelines and orchestrating multi-agent systems that solve actual problems, not just theoretical ones. I believe that mastery comes from teaching, which I practice by mentoring 50+ peers in my campus AI club. I am looking for a team where I can contribute to meaningful research while learning from world-class engineers.

EDUCATION

Indian Institute of Information Technology, Raichur   Karnataka, India
Bachelor of Technology in Computer Science and Engineering; CGPA: 8.2/10   *Aug 2023 – July 2027*
– **Relevant Coursework:** Deep Learning, Natural Language Processing, Artificial Intelligence, Probability & Statistics, Data Structures

TECHNICAL SKILLS

Languages: Python (Advanced), SQL, C++
AI & ML: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers, LLMs, Computer Vision
GenAI & Orchestration: LangChain, LangGraph, RAG, QLoRA Quantization, Agentic Workflows, Llama-3, Unsloth
Data & Databases: Pandas, NumPy, Milvus, ChromaDB, PostgreSQL, SQL
Deployment & MLOps: Docker, FastAPI, Groq LPUs, AWS (EC2, S3), Git, JMeter

EXPERIENCE

Coordinator & Lead Mentor   Dec 2024 – Present
DeepLabs (AI/ML Club), IIIT Raichur   *Raichur, Karnataka*
– Led technical workshops on Agentic AI and RAG, upskilling 50+ club members in cutting-edge AI technologies
– Mentored student teams for national hackathons, establishing MLOps and system design documentation standards
– Organized coding competitions and technical sessions, increasing peer engagement and problem-solving proficiency by 40% over one semester

PROJECTS

VitalSync-AI | *Medical Triage Chatbot & Optimization*   ⌾ *GitHub*
– Engineered an evidence-based diagnostic system using Llama-3 8B and RAG, delivering precise medical triage responses
– Reduced model memory footprint by 75% through 4-bit QLoRA quantization, enabling inference on consumer-grade hardware
– Integrated IBM WatsonX and Hugging Face Transformers to implement hallucination detection and fact-verification checks, ensuring trustworthy and accurate responses

SentinelAI-AML | *Multi-Agent Financial Security System*   ⌾ *GitHub*
– Designed a scalable Anti-Money Laundering platform using multi-agent workflows with LangGraph to automate fraud detection
– Increased transaction throughput by 3x (verified via JMeter benchmarking) through asynchronous FastAPI and Redis caching
– Containerized the solution with Docker and implemented health checks, achieving 99.7% availability during stress testing

NaturalSQL-AI | *Generative Database Assistant*   ⌾ *GitHub*
– Built a full-stack text-to-SQL system converting natural language queries into executable SQL with schema-aware prompting
– Optimized inference pipeline to achieve 5x faster query generation, reaching 400+ tokens/second via Groq LPU integration
– Deployed autonomous LangChain agents with database introspection for accurate, database-agnostic query generation

CERTIFICATIONS & ACHIEVEMENTS

Certifications: AWS Certified Cloud Practitioner | AI Engineer Agentic Track (Udemy) | Complete Generative AI with LangChain & Hugging Face (Udemy) | Machine Learning A-Z: Hands-On Python & R (Udemy)
Competitive Programming: LeetCode Top 6% Global Ranking (1800+ rating)
Hackathons: Flipkart GRiD 7.0 Shortlisted | Adobe India Hackathon Round 1 Qualifier | HP Power Lab 2.0 Participant