Syed Omer Shah

syedomer@buffalo.edu | (716)-290-9800 | linkedin.com/syed-omer-shah | github.com/Syedomershah99

EXPERIENCE

DEEPLEARNING.AI | Mentor, GenAI & LLMs
Jan 2025 - Present | Remote
- Achieved a resolution rate of **92%+** while mentoring 7 cohorts and supporting **280+** learners through the forum.
- Open-source contributor for productionized lab workflows, reducing failure cases and improving project completion consistency.

JK TECH | AI Intern
Mar 2025 - May 2025 | Remote
- Built an AI agent that parses resumes, ranks candidates, and auto-drafts outreach: $9.5\times$ lower latency, **+35%** accuracy; optimized pipeline for **800 resumes/sec**.
- Reduced manual ops by **95%** integrating CATSone & MongoDB ETL.

GNYAN.AI | Software Engineer Intern
Dec 2023 - May 2024 | Remote
- Quantized **4+** LLMs (Llama 3, Mistral 8×7B) with RAG + memory chaining, increasing answer relevance by **35%** F1 vs baseline
- Implemented Django/FastAPI auth servicing **10k+** req/mo. Deployed Ollama + Qdrant for **2×** faster semantic search in enterprise apps

RESEARCH

STATE UNIVERSITY OF NEW YORK RF | Research Aide
Sep 2025 - Present | Buffalo, NY

- Working with **Dr. Rutao Yao** in the Department of Nuclear Medicine on **SC-SPECT**; Built high-throughput 3D reconstruction workflows on HPC & Slurm; **30%** faster experiment cycles.

R & D CELL, MJCET | Head Undergraduate Researcher
Mar 2025 - Jun 2025 | Hyderabad, IN

- Developed a Mars Rover Prototype employing Autonomy & perception stack (VLM + Jetson Orin) with cloud telemetry; integrated AI pipeline for STT and TTS

PROJECTS

MEDGEMMA 270M | SLM for Radiology | **link**
- Fine-tuned Gemma 3 (270M) on MIRIAD using Unsloth FastModel for faster & lower VRAM training; includes Chat template formatting for supervised finetuning (SFT)

MEDILENS | Vision-Language QA | **link**
- Bilingual VLM trained on **1.6M+** samples; accuracy **+9%** EN/**+20%** AR vs. GPT-4 baseline

PUBLICATIONS

[1] **VMatter: AI-Enabled IoT Patient Monitoring & Room Automation**. Oral at ICMLDSSIA-2025; to appear in Springer Nature *Information Systems Engineering and Management*.

[2] **Declining ICU Mortality of COVID-19: A Multi-Center Study**. *Journal of Clinical Medicine Research*, Mar 2021.

EDUCATION

STATE UNIVERSITY OF NEW YORK AT BUFFALO
M.S. in Artificial Intelligence
Expected Jun 2027 | Buffalo, NY

SKILLS

PROGRAMMING
- Python • C++ • Java • Next.js
- JavaScript/TypeScript • Linux • Git

AI/ML & FRAMEWORKS
- PyTorch • TensorFlow • NumPy
- Pandas • SciPy • scikit-learn
- Transformers • Diffusion Models
- GRPO • PEFT • LoRA • QLoRA
- DeepSpeed • CUDA • cuDNN
- vLLM • ONNX • LangGraph

RETRIEVAL & RANKING
- RAG • FAISS • Qdrant • Pinecone
- Chroma • Elasticsearch • BM25
- Embeddings • Recommenders

MLOPS & RELIABILITY
- Kubernetes • Docker • CI/CD
- GitHub Actions • MLflow

COURSEWORK

GRADUATE
Analysis of Algorithm, Machine Learning, Artificial Intelligence, Reinforcement Learning, NLP, Blockchain, Deep Learning, Data Intensive Computing

UNDERGRADUATE
Python Programming, Probability Theory, Computer Architecture, Internet of Things, Neural Networks, Computer Vision and Pattern Recognition

CERTIFICATIONS
- Oracle Cloud Infrastructure Gen AI Professional (1Z0-1127-25)
- AWS Certified Cloud Practitioner (CLF-C02, 2025)

EXTRACURRICULAR
- 2025: "Deliver Results" Award Winner (AWS APJC, re:Invent)
- 2025 AWS Community Builder - AI Engineering
- 2024 Lead Organizer - DataNyx'24 (First 24-hr Datathon in Telangana)
- 2024 Founder - AWS Cloud Club (2000+ member community)