Nagur Shareef Shaik

404-203-9276 | shaiknagurshareef6@gmail.com | linkedin.com/in/nagur-shareef-shaik | github.com/ShaikNagurShareef

Summary

With 4.5+ years of expertise in Software Engineering & Machine Learning, I specialize in advancing multi-modal applications by customizing and fine-tuning LLMs and architecting autonomous Agentic AI systems. My proficiency in model development, optimization, and cutting-edge research, spanning 21 papers on Deep Learning Applications, sets me apart. I have a proven track record in developing scalable microservices & AI applications, successfully transforming theoretical insights into innovative, real-world solutions.

Education

Georgia State University — Atlanta, GA
Master of Science (Thesis) in Computer Science | GPA: 4.17/4.30 | 100% Scholarship — *Aug 2023 – May 2025*
Coursework: *Advanced Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Digital Image Processing, Computational Intelligence, Data Science*

Vignan's Foundation for Science, Technology & Research University — Guntur, AP
Bachelors of Technology (Honours) in Computer Science & Engineering | GPA 3.92/4.0 — *Jun. 2016 – May 2020*
- *University Gold Medalist and Best Outgoing Student of Computer Science Department*

Experience

UST Global Inc. — July 2024 – Present
Associate III, Data Science | GenAI — *(Remote) Aliso Viejo, CA*
- Architected **GenAI driven code modernization** solution to migrate complex legacy systems (.NET4 to .NET8; COBOL to React/Java Spring Boot) from monolithic to scalable, maintainable microservice architectures.
- Engineered a **LangGraph-based Agentic Framework** to generate **95%** functional code by auto-scaffolding enterprise-grade NX Monorepos (React Native, Nest.js), accelerating development by 70% at 60% cost.
- Pioneered advanced data solutions by collaborating with **Stanford AI Lab** on a **Text-to-SQL** framework (benchmarked on **BIRD**) and developing an **NL2SQL AgenticRAG** POC saving **4 FTEs** annually.
- Lead a 10-person AI team to architect **CodeCrafter**, an Agentic AI module for UST Aplha AI platform, securing **$18 million** in client contracts by delivering impactful **POCs**.

Georgia State University — Sep 2023 – May 2025
Graduate Research Assistant | TReNDS Center — *Atlanta, GA*
- Customized **Multi-modal LLMs** for automated report generation from medical images and text using **PyTorch**, achieving a **13.4% higher** BLEU4 than VisionGPT, and reducing inference time to **1.6 sec** per image.
- Developed image classification models using **TensorFlow/Keras**, for diagnosing chronic diseases like retinopathy, schizophrenia, breast cancer, and colon cancer from respective medical images, reducing **5%–7%** false negatives.

Carelon Global Solutions — Sep 2022 – Aug 2023
Software Engineer III | Elevance Health — *Hyderabad, TS*
- Developed **REST APIs** for **CO**mpensation **IN**centive **S**ystem, a **Microservices** based application, to validate, compute, and expedite incentive payments, achieving a **10%** reduction in processing time & enhancing scalability.
- Developed **Python** and **SQL** data clean-up scripts to resolve commission payment inconsistencies in health insurance policies, preventing over **$500K** in overpayments and streamlining processing time.
- Resolved critical production issues in commissions calculation flow, ensuring accurate agent advance computation, saving **$1.5 million** in historical commission overpayments. Received **Go Above Impact Award** for this work.

Tata Consultancy Services — Aug 2020 – Sep 2022
Systems Engineer | Analytics & Insights Business Unit — *Hyderabad, TS*
- Designed an Bug Root Cause Prediction System based on logs, by implementing **Attention LSTM** model in **Azure ML Studio**, cutting the debugging efforts and saving **3 full-time equivalents** annually.
- Implemented an **Azure DevOps** Model Deployment pipeline, reducing deployment time **from 2 to 1.2 hours** and increasing system availability by **25%**. Recognized as **Star Performer** of the team for this significant work.
- Built a custom SonarQube plugin for static code analysis, integrated into pre-build pipelines, reducing errors and vulnerabilities by **15%**, saving **6 hours** of manual review time per week, and adhering to coding standards.

Skills

Languages: Python, Java, SQL
Technologies: LangGraph, LangChain, Flask, FastAPI, Spring Boot, REST APIs, Microservices
Libraries: PyTorch, torchtune, TensorFlow, Keras, Scikit-Learn, OpenCV, NLTK, NumPy, Pandas, Matplotlib
Cloud & DevOps Tools: AWS (EC2, Lambda, SageMaker), Azure (AI Services, DevOps), CI/CD, Git, Docker
Research Interests: Deep learning, Multi-modal Learning, Attention Networks, Large Language Models, AgenticAI
Certifications: Azure AI Fundamentals, Deep Learning Specialization, Python for Everybody

Projects

InsuCompass | Github | Demo | *Python, LangGraph, Streamlit, Groq, ChromaDB* May 2025

- Developed an **Agentic RAG** system designed to demystify U.S. health insurance by ingesting data from official sources (CMS.gov, VA.gov) and delivering personalized plan recommendations in conversational style.
- Orchestrated a multi-agent workflow using **LangGraph** to deliver context-aware insurance advice, leveraging the **Groq API** for high-speed **LlaMA 3** inference & a local **ChromaDB** vector store for private user experience.

ScholarPulse | Github | Demo | *Python, LangChain, Streamlit, ChromaDB, SentenceTransformers* April 2025

- Architected an AI-powered research assistant based on **Advanced RAG** to analyze complex academic papers, delivering tailored question-answering, summarization, and code generation for users with diverse backgrounds.
- Engineered a **LangChain** workflow with advanced prompt engineering for **LlaMA-4 Maverick** insights and **Qwen-2.5** code generation, in a **multi-stage RAG pipeline**, cutting research comprehension time by **50%**.

Retinal Health Diagnostics | Github | Demo | *Python, TensorFlow, FastAPI, Streamlit* May 2024

- Developed an AI system with a novel attention mechanism that leverages global context to learn localized lesion-specific features for diagnosing chronic retinal diseases.
- Achieved state-of-the-art performance with accuracies of 97.5% for cataracts, 85.6% for diabetic retinopathy, and 94.6% for macular edema, highlighting strong clinical relevance.

Birthday Greetings App | Github | *Java, Spring Boot, JSP, MySQL, HTML, CSS, JS* April 2019

- In this project, we designed a Web Application that facilitates the end users to convey their wishes to friends by sending a greeting card to their email.

Master's Thesis

Attentive Multi-modal Learning for Medical Image Analysis | Advisor: Dong Hye Ye Aug. 2023 - May. 2025

- Integrating multi-modal data - such as medical images, clinical text, and genetics - through scalable attention networks and vision-language models has the potential to significantly enhance diagnostic accuracy and decision-making in schizophrenia diagnosis and medical report generation. By advancing deep learning techniques for multi-modal learning and focusing on explainable AI systems, this research aims to create efficient, robust, and responsible AI solutions that improve diagnostic precision and automate medical report generation, all while ensuring deployment in resource-constrained environments

Research Accomplishments

- Published **15** research papers in reputed journals and presented **7** papers at top conferences, including **ICIP 2024**, **ISBI 2024/25**, **ICASSP 2025**, and **MICCAI 2025** with **1K+** citations and **12** H-index, pioneering advancements in Attention models, Multi-modal learning, Transformers, and Vision-Language Models (Google Scholar).
- Active reviewer for **20+ journals** and conferences, evaluating **40+ articles** in Machine Learning.