Mufakir Qamar Ansari

Canton, MI | +1 (409) 827-9111 | mufakirqamaransari@gmail.com | Open to Relocate
LinkedIn | Google Scholar | Website

SKILLS & TOOLS

Proficient: Python, SQL, PySpark, PyTorch, Scikit-learn, XGBoost, PostgreSQL, Machine Learning, Deep Learning, Predictive Modeling, ETL Pipelines, Statistical Analysis

Familiar: AWS (SageMaker, S3), Azure, Databricks, Docker, Apache Airflow, FastAPI, Large Language Models (LLMs), Natural Language Processing (NLP), MLOps, Git

Exposure: Rust, CUDA, High-Performance Computing (SLURM), Neo4j, Qdrant, Time Series Forecasting, Causal Inference

RESEARCH EXPERIENCE

Wright State University, Dayton, OH (Jan 2026 – Present)
Research Assistant

- Architected a scalable 19-stage HPC ETL pipeline on the OSC supercomputer (SLURM) for Ebola RNA-Seq analysis, processing 356 samples and 15M+ read pairs with reproducibility guaranteed via AWS S3 and Azure Databricks.
- Developed the DDFF (Dynamic Data Feature Framework), engineering a robust feature selection pipeline to strictly optimize dataset statistics and classifier interpretations for high-dimensional data processing.

The University of Toledo, Toledo, OH (Aug 2023 – Apr 2025)
Research Assistant, Transportation Systems Research Lab & High-Performance Computing Lab

- Processed and modeled 37,000+ complex geospatial records, engineering full-stack ensemble models (Random Forest, XGBoost) that achieved state-of-the-art predictive accuracy for the Ohio Department of Transportation (FHWA/OH-2025/19).
- Slashed deep learning training time by 50% (2x speedup) and accelerated GPU execution by 8.5% via distributed HPC workload optimization, while engineering a predictive scheduling model that reduced overall carbon footprint by 5%.

PROFESSIONAL EXPERIENCE

Orcinus IT Solutions Private Limited, Srinagar, India (May 2018 - Dec 2022)
Technical Lead

- Accelerated data pipeline execution by 35% across 7 high-volume SaaS enterprise clients via scalable ETL workflows (Apache Airflow, PySpark, Azure), and boosted model reliability by 30% through a comprehensive MLOps framework.

MyFajir IT Solutions Private Limited, Srinagar, India (Sep 2017 - Mar 2020)
Senior Engineer, then Consultant

- Engineered production-grade RESTful APIs (Flask, FastAPI) maintaining 99.9% uptime for 800+ daily analytics transactions, and slashed client operational overhead by 20% by deploying a robust Azure cloud ERP platform.

PROJECTS

High-Performance Computing Pipeline for Ebola RNA-Seq Analysis

- Architected an end-to-end, 19-stage ETL pipeline on the OSC supercomputer via SLURM array parallelization, processing 356 samples and over 15 million read pairs from the 2014 West African Ebola outbreak.
- Engineered robust dual-quantification workflows (HISAT2, Kallisto) and orchestration logic, ensuring seamless terabyte-scale data processing and generating cross-validated transcriptomic analytics.

AI for Science: Clinical-Grade Cancer Detection via Deep Learning

- Achieved clinical-grade cancer detection (AUC-ROC = 0.950, AUC-PR = 0.886), outperforming ImageNet transfer learning baselines by +0.032 and +0.044 respectively, by architecting a domain-specific deep learning pipeline across 277,500+ Invasive Ductal Carcinoma image patches.
- Enhanced clinical interpretability by pioneering representation learning workflows via self-supervised learning (SimCLR), utilizing t-SNE, UMAP, and Grad-CAM visualizations to validate model attention to critical histopathological markers.

Privacy-Preserving AI Agent & Knowledge Graph Engine

- Engineered a highly secure, cross-platform AI agent ecosystem (Tauri/Rust, FastAPI, Next.js) featuring robust AES-256-GCM encryption and local LLM inference, guaranteeing strict data sovereignty.
- Architected a real-time Neo4j Knowledge Graph powered by custom NLP pipelines to continuously ingest, map, and analyze complex cross-platform user activity (browser engagement, desktop context) into deeply personalized insights.

SELECTED PUBLICATIONS

- M. Q. Ansari, et al., "From Text to Returns: Using Large Language Models for Mutual Fund Portfolio Optimization and Risk-Adjusted Allocation", Submitted, The Global AI Summit 2025.
- M. Q. Ansari, et al., "High-Sensitivity Detection of Invasive Ductal Carcinoma via Domain-Specific SimCLR Pre-Training," Under Review, Signal, Image and Video Processing.

EDUCATION

Master of Science in Computer Science & Engineering (Aug 2023 – Aug 2025)
The University of Toledo, Ohio, USA — *Track: Artificial Intelligence* | *GPA: 3.91/4.00*

Bachelor of Technology in Electronics & Communication Engineering (Jul 2009 – Jul 2013)
National Institute of Technology, Srinagar, India