Mufakir Qamar Ansari

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

Summary

Ph.D. Researcher in Artificial Intelligence specializing in AI for Science, representation learning, and high-performance computing. Proven track record of architecting scalable ML pipelines, distributed GPU training, and privacy-preserving RAG systems. Passionate about leveraging large language models and self-supervised learning to accelerate scientific discovery and build intelligent, safe AI agents.

Skills & Tools

Programming: Python, Rust, SQL, PySpark, CUDA, C++, Java
Data Science & ML: PyTorch, TensorFlow, Scikit-learn, Large Language Models (LLMs), RAG, Self-Supervised Learning, LangChain, Computer Vision, NLP
Databases: PostgreSQL, MySQL, MongoDB, Vector Databases (ChromaDB), Graph Databases (Neo4j), Knowledge Graphs
Statistics & Modeling: A/B Testing, Causal Inference, Time Series Analysis, Optimization Methods
Cloud & MLOps: AWS (SageMaker, S3), Azure (Databricks), GCP, Docker, MLflow, Apache Airflow, Git

Professional Experience

Orcinus IT Solutions Private Limited, *Srinagar, India* *May 2018 - Dec 2022*
Technical Lead
- Directed stakeholder engagements for seven SaaS enterprise clients, designing and deploying end-to-end ML solutions and architecting scalable ETL pipelines (Airflow, PySpark, Azure) that reduced latency by up to 35%.
- Orchestrated a comprehensive MLOps framework ensuring rigorous version control, monitoring, and governance for production systems, resulting in a 30% improvement in model reliability across deployments.

MyFajir IT Solutions Private Limited, *Srinagar, India*
Senior Engineer, then Consultant *Sep 2017 - Mar 2020*
- Architected and deployed production-grade RESTful APIs using Python frameworks (Flask/FastAPI) for real-time analytics integration, processing 800+ daily transactions with 99.9% uptime; built interactive web dashboards for cross-functional stakeholder decision support.
- Collaborated in agile development environment using JIRA for sprint planning and iterative delivery of ML solutions; engineered cloud-based ERP platform (Azure) reducing client operational overhead by 20%.

PanunKart.com, *Srinagar, India* *Aug 2013 to Sep 2017*
Business Operations Lead
- Designed and implemented data strategy that drove 150% sales growth through A/B testing and market segmentation analysis. Enhanced data-driven decision-making capabilities resulting in improved business intelligence.

Research Experience

Wright State University & The University of Toledo, *Dayton & Toledo, Ohio* *Jun 2024 – Present*
Research Assistant, Transportation Systems Research Lab
- Architected and deployed full-stack machine learning solutions utilizing ensemble methods (Random Forest, XGBoost, LightGBM, Neural Networks) to process 37K+ complex records, achieving state-of-the-art predictive accuracy.
- Engineered production-grade, terabyte-scale data pipelines utilizing AWS (S3, EC2, SageMaker) and Azure Databricks; authored comprehensive technical runbooks to ensure reproducibility.

The University of Toledo, *Toledo, Ohio* *Aug 2023 – May 2024*
Research Assistant, High-Performance Computing Lab
- Spearheaded the optimization of distributed GPU-accelerated ML workloads on HPC clusters, improving kernel execution speed by 8.5% and reducing model training time by 2x—skills directly applicable to scaling foundation models.
- Engineered a Python-based predictive model correlating HPC energy consumption with grid emissions, driving smarter workload scheduling and achieving a 5% reduction in overall carbon footprint.

Education

Doctor of Philosophy in Computer Science & Engineering *Jan 2026 – Present*
Wright State University, Dayton, USA
Master of Science in Computer Science & Engineering *Aug 2023 – Aug 2025*
Track: Artificial Intelligence *GPA: 3.91/4.00*
The University of Toledo, Ohio, USA
Bachelor of Technology in Electronics & Communication Engineering *Jul 2009 – Jul 2013*
National Institute of Technology, Srinagar, India

Projects

- **AI for Science: Clinical-Grade Cancer Detection via Deep Learning [Python, PyTorch, ResNet50, SimCLR]**[GitHub]
  - Architected a domain-specific deep learning pipeline for Invasive Ductal Carcinoma (IDC) detection across 277K+ pathology image patches, achieving clinical-grade performance (AUC-ROC = 0.950, FNR = 0.34%).
  - Engineered representation learning workflows using self-supervised learning (SimCLR), enhancing clinical interpretability via t-SNE, UMAP, and Grad-CAM visualizations.
- **Privacy-Preserving AI Agent & Knowledge Graph Engine [Python, Rust, Local LLMs, RAG, ChromaDB]**[GitHub]
  - Spearheaded the development of a local-first AI intelligence platform that captures multimodal digital signals to construct dynamic, real-time knowledge graphs of user behaviors and workflows.
  - Engineered secure, on-device inference engines utilizing local LLMs and Retrieval-Augmented Generation (RAG) to synthesize personalized insights without compromising user data privacy.
  - Architected a zero-knowledge privacy framework that cryptographically separates enterprise and personal data flows while maintaining real-time contextual reasoning capabilities.
- **Production LLM & RAG Implementation: Hybrid Optimization System[LangChain, Hugging Face, Vector DBs]**[GitHub]
  - Orchestrated an end-to-end LLM system featuring a RAG pipeline and ChromaDB vector database for complex financial data processing and entity extraction.
  - Evaluated and benchmarked multiple foundation models (Phi 2, Mistral 7B, Zephyr 7B) by optimizing prompt engineering, vector search latency, and production monitoring protocols.
  - Synthesized production-ready LLM inference patterns, achieving superior risk-adjusted returns and robust performance stability.

Research Papers

- **M. Q. Ansari**, et al., "Dy-Part: A Dynamic, Noise-Aware Scheduler for Optimizing Hybrid Quantum-Classical Algorithms," *Under Review, Quantum Machine Intelligence.* Preprint available: DOI:10.21203/rs.3.rs-8041248/v1.
- **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.* Preprint available: DOI: arXiv:2512.05907v1.
- **Ansari, M. Q.** "A Comparative Analysis of Deep Learning Architectures for Probabilistic Electricity Price Forecasting." *Under Review, Engineering Applications of Artificial Intelligence*, (Submitted Aug 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.* Preprint available: DOI:10.21203/rs.3.rs-8031909/v1.
- E. Chou, R. Rai, **M. Ansari**, S. Khanal, A. Davenport, "From Data to Action: Leveraging Machine Learning/Artificial Intelligence to Guide Proactive Pedestrian and Cyclist Safety Initiatives," *Ohio Department of Transportation Final Report FHWA/OH-2025/19*, August 2025. Available: SJN 136837.
- Hasan, T., Hossain, A., **Ansari, M. Q.**, & Syed, T. H. (2025). "Enhanced Intrusion Detection in IIoT Networks: A Lightweight Approach with Autoencoder-Based Feature Learning." *Proceedings of the 10th International Conference on Internet of Things, Big Data and Security (IoTBDS 2025)*, pp. 207-214.
- **M. Q. Ansari**, et al., "Racing to Idle: Energy Efficiency of Matrix Multiplication on Heterogeneous CPU and GPU Architectures," *American Journal of Computer Science and Technology (AJCST),* ISSN Online: 2640-012X; ISSN Print: 2640-0111, 2025. Preprint available: DOI:10.21203/rs.3.rs-7890483/v1.
- **M. Q. Ansari**, et al., "Accelerating Matrix Multiplication: A Performance Comparison Between Multi-Core CPU and GPU", Preprint available: DOI:10.48550/arXiv.2507.19723.

Certifications

SQL for Data Science	*University of California, Davis*
Data Science Math Skills	*Duke University*
Introduction to Statistics	*Stanford University*
Distributed Computing with Spark SQL	*University of California, Davis*
Business Metrics for Data-Driven Companies	*Duke University*
Managing Machine Learning Projects	*Duke University*
Data Visualization and Communication with Tableau	*Duke University*