

ABRAR ZAHIN

650 E Tyler Mall Tempe, Arizona 85281

📞 435-764-1745 📩 abrarzahin303@gmail.com 💬 [LinkedIn](#) 🌐 [Website](#)

Personal Summary

Machine learning (ML) researcher specializing in multimodal foundation models for computational pathology, integrating image, text, and graph-structured data. Experienced in developing scalable and interpretable statistical ML models for large-scale networked systems, including probabilistic graphical models and generative methods. Strong background in computational statistics, deep learning, and large-scale inference, with a focus on building transferable representations and practical ML systems for real-world scientific applications.

Skills Summary

Languages: Python, SQL, R, MATLAB, C/C++, SAS, SPSS | **ETL:** Apache Airflow, Apache Kafka, PySpark

Frameworks and Libraries: PyTorch, TensorFlow, JAX, Keras, Pandas, NumPy, Scikit, NLTK, OpenCV, Matplotlib

DevOps and Cloud Development: CI/CD (Azure DevOps, GitHub Actions), Git, FastAPI, REST APIs, AWS, GCP, Docker

LLM/GenAI Development: Hugging Face Transformers, LangChain, LangGraph, Autogen, Pydantic AI, Vector Embeddings (Nomic), Vector Databases (Chroma, Supabase), GenAI APIs (OpenAI, Groq).

Research & Technical Expertise: Deep Learning, Statistical Modeling, Bayesian Inference, Causal Inference, Generative Models (VAEs, Diffusion Models), Multimodal Models, Recommendation Systems, Optimization Algorithms

Professional & Educational Background

Mayo Clinic: Machine Learning Research Fellow	Jan 2026 - Current
Mayo Clinic: Machine Learning Research Intern	Aug 2025 – Dec 2025
Arizona State University (ASU): PhD in Electrical Engineering	Jan 2020 – Dec 2025
Utah State University (USU): MSc in Electrical Engineering	Aug 2017 – Dec 2019
Energypac Engineering Ltd.: Data Scientist	Jan 2015 – Jul 2017

Research Papers

Robust Model Selection of Gaussian Graphical Models [Link] | Published in [TMLR](#)

- Developed a novel algorithm for network analysis that recovers complex structures in noisy data, enabling more robust and efficient optimization of recommendation systems, supply chains, and financial systems
- Our algorithm **transcends the fundamental limitations** of current algorithms in learning complex networks
- Implemented our algorithm on both simulated graphs and **real-world networks**

Rapid Change Localization in Gaussian Graphical Models [Link] | Published in [ICASSP](#)

- Developed a **novel** algorithm for rapid change localization in large-scale networked systems
- Computationally efficient and performs change localization with at least **20% faster than the baseline** algorithms
- Applicable to real-time system monitoring and anomaly detection in sensor networks

Computationally Efficient Active Learning of Gaussian Graphical Models [Link] | Published in [ASILOMER](#)

- A **novel** algorithm that achieves **exponential runtime reduction** for structure learning for large-scale networked systems
- Significantly reduced query complexity, enabling deployment in **low-resource or real-time environments**

Semi-supervised Learning of Fall Down Action [Link] | Published in [AICON](#)

- Developed a semi-supervised classifier with **Variational Autoencoder (VAE)** and **Convolutional Neural Network (CNN)**
- Our classifier is at least **5% more accurate** in classifying different fall down actions from a real-world data set

Efficient Smart Health Monitoring of Large-scale Networks [Link] | Published in [IETC](#)

- Developed a **novel** algorithm for smart healthcare monitoring, primarily using **Denoising Autoencoder** and **CNN**
- Our algorithm** is at least **27% faster** than the current **state-of-the-art framework**

• Sample-Query Complexity Trade-Offs in Learning and Testing Tree-Structured Graphical Models | (*In preparation*, [ICML](#))

• Hypothesis Testing between Two Gaussian Graphical Models using Covariance Queries | (*In preparation*, [AISTATS](#))

Notable Projects

- Causal Inference with ML:** Implemented uplift modeling and causal discovery methods (EconML, DoWhy) to estimate treatment effects, improving decision-making strategies in observational datasets [Code]
- Summarization API:** Deployed a Hugging Face summarization model as a production-grade API using **FastAPI**, **Docker**, and **Kubernetes**, enabling autoscaling and high availability [Code]
- Diffusion Model API:** Built a full-stack text-to-image generation pipeline with a web frontend, REST API, and GPU-backed inference, orchestrated via **Docker** and **Kubernetes** [Code]
- RAG Pipeline:** Implemented an end-to-end retrieval-augmented generation system for custom PDFs using LLaMAIndex and vector embeddings, with real-time querying via **FastAPI** [Code]
- Streaming ETL:** Built a real-time pipeline with **Kafka**, **Airflow**, and **Parquet**, using Bronze–Silver–Gold layers, DQ checks, and SLA monitoring; achieved **99% DAG success** with data freshness under **60s** [Code]
- Certificates and Notable Technical Reports : [Gen AI with LLMs](#), [Data Analysis with SQL](#), [Streaming ETL](#), [Link 1](#), and [Link 2](#).