Abrar Zahin

650 E Tyler Mall Tempe, Arizona 85281
435-764-1745 abrarzahin303@gmail.com LinkedIn Website

Personal Summary

Statistical machine learning researcher with expertise in generative AI, multimodal learning, and large-scale model optimization. My research focuses on diffusion models, 3D and video generation, and retrieval-augmented systems, with a strong foundation in probabilistic modeling and efficient inference techniques. I bring experience developing scalable ML pipelines using PyTorch, leveraging high-performance computing, and optimizing models for deployment. Passionate about connecting foundational research to real-world applications in controllable content synthesis and embodied AI.

Skills Summary

Languages: Python, MATLAB, SQL, SPSS, SAS, R, C, and C++

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

Data Visualization and Analysis: Tableau, Microsoft PowerBI, Seaborn, GraphPad

High-Performance Computing: Batch Scripting, Python Multi-Processing, GPU Clusters, Cython

Cloud Development: AWS, GCP, Docker, MySQL

DevOps: CI/CD (Azure DevOps, GitHub Actions, Jenkins), Git, FastAPI, Flask, REST APIs

Research & Technical Expertise: Statistical Machine Learning, Generative AI (Diffusion Models, GANs), Multimodal Learning (Vision-Language, 3D, Video), Retrieval-Augmented Generation (RAG), Probabilistic Modeling & Inference, Causal Learning & Model Interpretability, and Embodied AI and Reinforcement Learning

Education

Arizona State University (ASU): PhD in Electrical Engineering — Jan 2020 – Current

Utah State University (USU): MSc in Electrical Engineering — Aug 2017 – Dec 2019

Research Papers (Published)

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 **provably low latency** 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

- Developed a **novel** computationally efficient 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**

Notable Projects

- **Image Denoising Convoulutional Autonencoder:** Designed and trained a convolutional **autoencoder** for image denoising, achieving a **3dB PSNR improvement** over baseline models using Gaussian noise.[Code].
- **Summarization API**: Deployed a Hugging Face summarization model as a production-grade microservice using **FastAPI** (API development), **Docker** (containerization), and **Kubernetes** (orchestration) with autoscaling, load balancing, and rolling updates for high availability. [Code] [Report]
- **Diffusion Model API:** Deployed a text-to-image **diffusion model** as a full-stack ML application using **FastAPI** for serving, **Docker** for containerization, and **Kubernetes** for scalable orchestration, integrating a web frontend, REST API, and GPU inference backend for end-to-end generative image synthesis. [Code] [Report]
- Predicted electric vehicle charging demand using a **Long Short Term Memory**, achieving a **20% reduction in RMSE** over baseline time-series models, enabling smarter grid load balancing for EV infrastructure.[Code].
- **Built an end-to-end Retrieval-Augmented Generation (RAG) pipeline** using LLaMAIndex and local embedding models for question answering over custom PDFs, and **deployed a FastAPI interface** supporting real-time queries with **chunked document parsing**, **vector indexing**, and **LLM-based response generation**. [Code].
- **GAN** for speech generation and **Transformer** for speech anonymization [Code].
- Notable Technical Reports: Link 1, Link 2, Link 3, and Link 4.