Md Shakhrul Iman Siam

Columbus, Ohio, USA
siam.5@osu.edu | +1 (614) 205 9451
shakhrul-iman-siam | shakhrulsiam268 | sites.google.com/view/shakhrul-iman-siam

Research Interest

I am a second-year Ph.D. Student specializing in Computer Science with a focus on Multimodal Foundation Model and Generative AI. My research interests include building multi-modal LLM for **Healthcare applications**, **Retrieval-Augmented Generation**, **Automated Speech Recognition.**

Education

- **The Ohio State University** — Columbus, Ohio, USA
  Ph.D. in Computer Science and Engineering (CSE) — August 2023 - 2028 (Expected)
  Selected Courses: Artificial Intelligence, Speech and Language Processing, Neural Networks
  GPA: 4.0/ 4.0

- **Bangladesh University of Engineering and Technology (BUET)** — Dhaka, Bangladesh
  BSc. in Electrical and Electronic Engineering (EEE) — Feb 2016 - Feb 2021
  Selected Courses: Digital Signal processing, Microcontroller, Wireless communication

Work Experience

Research Intern — USA
- Microsoft — June 2025 - present
  Research Focus: Automated Speech Recognition

Graduate Research Associate, The Ohio State University — Columbus, Ohio, USA
- OSU AIoT and Machine Learning Systems Lab — August 2023 - Present
  Advisor: Prof. Mi Zhang
  - **LLM-based Medical Question-Answering Assistant using Retrieval-Augmented Generation**
    - Developed a Question-Answering Assistant capable of Medical Domain QA tasks.
    - Integrate various RAG techniques like Naive RAG, Graph RAG, and Light RAG.
    - Designed a knowledge graph and implemented Vector Database using FAISS index and Pinecone.
    - Implement various pre-retrieval optimizations for efficient similarity searching.
  - **Reading Recognition in the Wild**
    Submitted to ICCV 2025
    - Collected a large-scale multi-modal reading dataset of 100 hours of egocentric videos using Meta Aria Glass.
    - Develop a transformer-based model that detects whether and when a user is reading using different input modalities (egocentric video, eye gaze, IMU)
    - Developing a benchmark using both gaze information and RGB image data to classify reading and non-reading activities.
  - **Multimodal Foundation Model for Respiratory Health Diagnosis**
    - Developed a Multimodal foundation model that will capture sounds from a smartphone microphone and perform several downstream tasks like disease (Covid-19, TB, COPD) detection, generating health reports, and Question-answering.
    - Utilize LLaMA-3 with PEFT and LoRA techniques to fine-tune multimodal models from cough sounds and self-reported patient text metadata.

Machine Learning Engineer — Dhaka, Bangladesh
- ACI Limited (Full-time) — October 2021 - July 2023
  Domain Experience: Computer Vision, Speech Recognition, Predictive analysis, NLP
  - **Person re-identification and tracking using multiple CCTV camera feeds.**
    - Developed a system to re-identify, track, and monitor customers' movement in Retail outlets from CCTV camera network.
    - Person Detection using YOLOv5 model trained on a custom dataset, and Re-identification using embedding generated by a Resnet-50 model trained on triplet loss.
    - Developed a 2D projection algorithm to map each persons' position from camera image to floorplan.
    - Developed a novel location embedding-based tracking algorithm to improve the re-identification accuracy.

- **Facial Recognition based automatic attendance system.**
  - Implemented face detection using Retinaface and face recognition using Arcface.
  - Converted the model to Tflite to improve inference speed by 4×.
  - Designed the database (MySQL) for daily attendance entry and built a dashboard with Django.
  - Currently deployed on multiple facilities under the same organization

- **Medicine name recognition system from handwritten prescription.**
  - Developed a ROI Extraction model by using LayoutLM.
  - Developed OCR models using EasyOCR, PyTesseract, and PaddleOCR to recognize medicine names.
  - Integrated FuzzyWuzzy, and CTC Beam Search for post-processing.
  - Designed Database (MySQL) and Built a Dashboard using Django.

- **Voice Controlled Conversational Chatbot.**
  - Implemented a streaming ASR model and a Text-to-speech model for automated speech recognition of users.
  - Designed a sliding window technique to provide the ASR model with real-time streaming capabilities.
  - Implemented RASA (a conversational AI framework) as a chatbot agent.
  - Built backend API services using Django & Django Rest Framework

Research and Development Executive Dhaka, Bangladesh

- Spectrum Engineering Consortium Limited (Full-time) April 2021 - Sep 2021

Domain Experience : Robotics, Computer Vision, Microcontroller, IoT

- Developed a prototype of a low-cost mobile device, that can assist a visually disabled person to read, recognize person, and detect objects using voice command.
- Worked with various Microcontroller devices (Arduino, Raspberry-pi, ATMEGA, Nvidia Jetson Nano), Wireless communication modules (NRF24L01, HC-05), Hybrid Stepper Motor and drivers.
- Designed Printed Circuit Board (PCB) for various applications.
- Worked on Control and Power system design of a Robotic arm.

Publications

Published

- **Shakhrul Iman Siam**, Hyunho Ahn, Li Liu, Samiul Alam, Hui Shen, Zhichao Cao, Ness Shroff, Bhaskar Krishnamachari, Mani Srivastava, and Mi Zhang. "Artificial Intelligence of Things: A Survey" ACM Transactions on Sensor Networks (2024).
- Chenning Li, Yidong Ren, Shuai Tong, **Shakhrul Iman Siam**, Mi Zhang, Jiliang Wang, Yunhao Liu, and Zhichao Cao. "ChirpTransformer: Versatile LoRa Encoding for Low-power Wide-area IoT." In Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services (Mobisys 2024).
- **Md Shakhrul Iman Siam**, and Subrata Biswas. "A Deep Learning Based Person Detection and Heatmap Generation Technique with a Multi-Camera System." In 2022 12th International Conference on Electrical and Computer Engineering (ICECE 2022).
- **Md Shakhrul Iman Siam**, Md Saiful Bari Siddiqui, Mushfiqul Abedin, and Mohammed Imamul Hassan Bhuiyan. "Bioradiolocation-Based Multi-Class Sleep Stage Classification Using Time and Frequency Features with Random Forest Classifier." In 2022 12th International Conference on Electrical and Computer Engineering (ICECE 2022).
- KM Naimul Hassan, Subrata Kumar Biswas, Md Shakil Anwar, **Md Shakhrul Iman Siam**, and Celia Shahnaz. "A Dual-Purpose Refreshable Braille Display Based on Real Time Object Detection and Optical Character Recognition." In 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON 2019).

Under Review

- Charig Yang, Samiul Alam, **Shakhrul Iman Siam**, Michael J. Proulx, Lambert Mathias, Kiran Somasundaram, James Fort, Omkar Parkhi, Yuheng Ren, Mi Zhang, Yuning Chai, Richard Newcombe, Hyo Jin Kim. "Reading Recognition in the Wild" (NeurIPS 2025).
- **Shakhrul Iman Siam**, Tiantian Feng, Jiankun Zhang, Shrikanth Narayanan and Mi Zhang. "RespiraMFM: A Multimodal Foundation Model with Contrastive Audio-Language Alignment for Respiratory Disease Identification." (EMNLP 2025).
- Hyunho Ahn, Hansol Lee, **Shakhrul Iman Siam**, Sachin Kumar. "How Much Context Is Enough? Evaluating the Role of Audio and Textual Context in ASR Systems." (EMNLP 2025).

Skills Summary

Programming Languages	Python, C, C++, JavaScript, MATLAB, Assembly(ARM), Unix, Shell Scripts
Machine Learning Domain	Computer Vision, Generative AI, Foundation Model, Face-Recognition, Predictive Analysis, Recommendation System, Chatbot, Reinforcement Learning
ML-DL Frameworks	Pytorch, TensorFlow, Keras, Scikit-learn, NLTK, Pandas, OpenCV, Numpy, Unity
LLM Frameworks	MLC-LLM, DeepSpeed
Backend Frameworks	Django, Django Rest Framework
Databases	MySQL, Oracle
Web Development	HTML, CSS, BootStrap
Cloud & Containerization	AWS, Docker, Kubernetes
Hardware & IoT	Arduino, Raspberry pi, Microcontroller, Meta Aria Glasses

Selected Projects

- **Generating Customer Heatmap of a retail outlet from CCTV camera :** Using Yolov5 object detection model, Homorgaphy transformation and Kernel density estimator to generate customer density map on a floorplan of a retail outlet using video taken from CCTV camera.
- **Blind vision Assistance :** Prototype of a low-cost mobile device, that can assist a visually disabled person to read, recognize person, and detect objects using voice command. There are three main features of this system- Object detection, Face Recognition, and optical character recognition.
- **Real-time vehicle detection and tracking at junction using a fisheye camera :** The Video and Image Processing cup 2020 challenge focuses on fisheye cameras mounted into street lamps at junctions and vehicle detection and tracking to be used for a junction management system to optimize the flow of traffic and synchronize with other junctions to obtain bottleneck performances throughout the city.
- **Open Source Contribution in developing a pypi package:** A Python library for offline reverse geocoding. Reverse Geocode BD takes a (latitude , longitude) coordinate and returns the Division, District, Upazila and Thana of any location in Bangladesh.
- **Unsupervised Synthetic/Fake Speech Detection.** Our research introduces a novel unsupervised model, which we have termed "Anocoder". This model employs a one-dimensional convolutional autoencoder and is designed to differentiate between authentic and synthetic or fabricated speech.
- **Web Scrapping and Sentiment Analysis from Social Media post:** Using Selenium for web scrapping and analysis of social media and online newspaper posts for sentiment analysis. Tech: Selenium, BeautifulSoup

Honors and Awards

- First Runner up at Robi **Datathon 2.0** (2022)
- 5th at **IEEE Video and Image Processing Cup** (2020)
- Champion of Bangladesh Section and World Finalist, **IEEE YESIST12 Innovation Challenge** (2019)
- 1st Runner up, BUET CSE Day **Math Olympiad** (2017)
- 1st Runner up, BAS Divisional **Science Olympiad** (2015)
- Champion , Dutch Bangla Bank Prothom Alo Regional **Math Olympiad** (2015)
- 2nd Runner Up , Dutch Bangla Bank Prothom Alo Regional **Math Olympiad** (2014)
- Champion , Dutch Bangla Bank Prothom Alo Regional **Math Olympiad** (2011)

Reference

Mi Zhang
Professor - Department of CSE, The Ohio State University
Columbus, Ohio, USA.
Email: mizhang.1@osu.edu