KMA Solaiman

Ph.D. Candidate
Department of Computer Science
Purdue University
West Lafayette, IN 47907

LWSN B132
+1 765-775-8230
https://ksolaiman.github.io/
ksolaima@purdue.edu

PERSONAL PROFILE

Ph.D. candidate with 6+ years of experience teaching courses in both undergraduate and graduate levels. Built systems with societal impacts. Mentored more than 12 undergraduates and Master's. Collaborated with multiple industry and academia partners. Published in 1 journal, 2 conferences, and 3 workshops, with 1 journal and 1 conference paper in preparation.

EDUCATION AND HONORS

Ph.D. in Computer Science July 2023
Purdue University, Department of Computer Science West Lafayette, IN

- Advisor: Bharat K. Bhargava
- Graduate School Summer Research Grant (2018)
- Research Area: Multimodal Information Retrieval and Open-world Data Management Systems

M.Sc. in Computer Science Dec 2022
Purdue University, Department of Computer Science West Lafayette, IN

- Cumulative GPA: 3.63/4.00
- Relevant Coursework: Crowd-sourcing and Social Computing, AI meets Sustainability, Machine Learning, Data Mining, Distributed Database Systems, Compiling And Programming Systems, Data Communication and Networks, Algorithms

B.Sc. in Computer Science and Engineering (CSE) July 2014
Bangladesh University of Engineering and Technology (BUET), Department of CSE Dhaka, Bangladesh

- 16^{th}/ 153, Dean's List, University Merit Scholarship, University Stipend Award
- Cumulative GPA: 3.79/4.00, In-major CGPA: 3.88/4.00
- Thesis: *Minimal Parameter Clustering of Complex Shaped and Different Sized Dataset with Noise and Outlier Handling*

TEACHING EXPERIENCE

Graduate Teaching Assistant
Purdue University West Lafayette, IN
Department of Computer Science Aug 2016 - Dec 2019, Aug 2022 - Present

- Undergraduate Courses: Problem Solving and Object-Oriented Programming (3 semesters), Data Structures (3 semesters), Introduction to Relational Database Systems (2 semesters)
- Graduate Courses: Simulation & Modeling of Computer Systems, Data Communication and Computer Networks
- Graded 2+ projects (with Big Data project for Database) each semester for full class and the exams and provided feedbacks
- Instructed in labs and PSOs with ∼30 students (from freshmen to graduate levels)
- Designed home work, assignments and exams
- Mentored students for final course projects in OOP and paper reproduction in Networks
- Course development for OOP and Simulation & Modeling

Guest Lecturer
Purdue University

- Situation Awareness, Adversarial ML, and Explainable AI
- Information Retrieval

Lecturer
Ahsanullah University of Science & Technology, *United International University* Dhaka, Bangladesh
Department of Computer Science & Engineering Aug 2014 - Jul 2016

- Lectures at the undergraduate level: Programming Language (106 students), Network Programming (143 students), Database (132 students), Software Development, Simulation and Modelling, Computer Graphics
- Conducting labs and supervising group projects in Database, Networking, and Software Engineering
- Student advising, participating in accreditation, and curriculum development for lecture and lab courses in CS
- Handled 12-18 credit hours each semester with academic services

SKILLS

- **Software Stacks**: Docker, Kafka, Kubernetes, Conda, Git (Version Control), Amazon Web Services (AWS)
- **Programming Languages**: Python, Java, C, C++, x86, Prolog, PHP, HTML, XML, CSS, JavaScript, Shell Programming
- **Deep Learning Frameworks**: PyTorch (*High*), Caffe, Matlab (*Moderate*)
- **Database and Big Data**: Oracle, Microsoft SQL Server, MySQL, PL/SQL, PostgreSQL, Spark, Hadoop, MongoDB
- **Data Science and Miscellaneous**: Data Science pipeline (cleaning, civilization, wrangling, visualization, modeling, interpretation), ETL, ERD, Hypothesis Testing, Google API, Jupyter Notebook

PUBLICATIONS

*[J: Journal, C: Conference, W: Workshop and Symposium, P: Posters]

**Co-first authors.

- **[J1]** **Kma Solaiman**, Tao Sun, Alina Nesen, Bharat Bhargava, and Michael Stonebraker. Applying Machine Learning and Data Fusion to the *Missing Person* Problem. **IEEE Computer**, *(Volume: 55, Issue: 6, June 2022)*.
- **[C3]** A. Nesen, **K. Solaiman** and B. Bhargava. Dataset Augmentation with Generated Novelties, **IEEE TransAI**, *2021*.
- **[W1]** S. Palacios and **K. Solaiman****, P. Angin, A. Nesen, B. Bhargava, Z. Collins, A. Sipser, M. Stonebraker. SKOD: A Framework for Situational Knowledge on Demand, In **POLY** at **VLDB**, *Springer 2019*.
- **[C1]** **Kma Solaiman**, MM Rahman, and N Shahriar. AVRA BANGLADESH: Collection, Analysis & Visualization of Road Accident Data in Bangladesh, **IEEE ICIEV**, *2013*.
- **[J2]** S. Islam and **K. Solaiman****, R. Oliveira, B. Bhargava. Domain Complexity Estimation for Distributed AI Systems in Open-World Perception Domain. **Artificial Intelligence** (Open-World AI), *July 2023*. [Submitted]
- **[C4]** **K. Solaiman** and B. Bhargava. Multi-modal Information Retrieval for Systems with Explicit Information Needs and Object Properties (FemmIR), In **SIGMOD** *2023*. [Submitted]
- **[W4]** **K. Solaiman** and B. Bhargava. Open-Learning Framework for Multi-modal Information Retrieval with Weakly Supervised Joint Embedding. In **AAAI Spring Symposium** on Designing Artificial Intelligence for Open Worlds, *2022*.
- **[W3]** **K. Solaiman** and B. Bhargava. Measurement of Novelty Difficulty in Monopoly. In **AAAI Spring Symposium** on Designing Artificial Intelligence for Open Worlds, *March 2022*.
- **[W2]** M. Stonebraker et al. Surveillance Video Querying With A Human-in-the-Loop, In **HILDA** with **SIGMOD**, *2020*.
- **[C2]** S Roy, **K Solaiman**, C Li, D Goldwasser. Identifying Bias in News Narratives Using Distant Supervision, **IWCS**, 2019. [Author in Previous Submissions]
- **[P1]** **Kma Solaiman**, AA Muzaddid. Minimal Parameter Clustering of Complex Shape Dataset with High Dimensional Dataset Compatibility, Presented in **BUET CSE Thesis Poster Presentation**, *2014*.

RESEARCH INTERESTS

Multi-modal Information Retrieval, Open-World AI, NLP+CV, and Data Management Systems.

RESEARCH EXPERIENCE

Research in Applications for Learning Machines (REALM). (Co-advised by: *Michael Stonebraker*)

- Proposed two *weakly supervised methods* for *multimodal information retrieval* based on graph matching techniques and representation learning approach. [C4, W4]
- In collaboration with local authorities, a *scalable cross-modal querying method* was built based on relational schema. The prototype was presented for finding a real-time scalable solution for *missing person search* where we used the real-world noisy and high dimensional data collected by the agencies. [J1]
- Proposed a novel *human attribute recognition model from unstructured text* using Word2Vec, SBERT sentence embedding and lexical database WordNet. [C4]
- Benchmarked attribute recognition models for videos and images along with the relational querying method. [C4]
- Scraped more than 6K tweets related to Cambridge, MA and used similarity search (e.g., LSI, LDA) to identify objects found in traffic videos. [W1]

Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON).

- Proposed domain complexity measures for distributed AI systems in perception domain and compared across different variations of MNIST. [J2]
- Proposed a graph state representation of the Monopoly environment for a reinforcement learning-based agent.
- Proposed estimation methods of difficulty levels of novelties introduced in Monopoly using the state representations. [W3]

Understanding Political Bias in News Articles using Social Media. [C2] *(Dan Goldwasser)*

- Experimented with Language Models e.g., Word2Vec, Glove, and Skip-Thought Vectors for representating news articles.
- Built *text classifiers for political bias* with SVM, Logistic Regression, LSTM, and Hierarchical LSTM.
- Designed joint representations for identifying political stance in news articles using weak supervision from tweets.
- Used Amazon MTurk to design a bias detection task and annotate newspaper articles with political bias.

Adversarial Attacks on Neural Networks. *(Miguel Villarreal Vasquez)* `Oct 2019`

- Experimented with a novel method for *tackling trojan attacks on deep neural network models*.
- Sampled a healing dataset from the LFW (Labeled Faces in the Wild) dataset and retrained the VGG-FACE model.

Unsupervised Learning [P1] *(Advisor:* Md. Monirul Islam*)* `Feb 2013 - Jul 2014`

- Proposed a *novel clustering algorithm for irregular* and *complex shaped data* with a single parameter, *filter-width*.
- Described an empirical method to dynamically find optimal value of *filter-width*.
- Extended the *Weka* framework to add the performance comparison of proposed algorithm with K-means, EM, etc.

Analysis & Visualization of Road Accident Data [C1] `Oct 2011 - May 2013`

- Implemented a novel web interface for collection of road accident data in Bangladesh after use case study.
- Performed *data analysis of road accidents* to compare and identify the prime contributors e.g., rural vs urban using Google APIs

Data Mining, Complex Network Analysis

- Implemented hand-gesture recognition from smart watch sensor data using Recurrent Networks (LSTM). (*He Wang*)
- Investigated *TribeFlow* for mining and predicting user preferences using hyperlink structure in Wikipedia.

COLLABORATIONS

Massachusetts Institute of Technology (MIT), University of Southern California (USC-ISI), Northrop Grumman Corporation (NGC), Institute for Defense Analyses (IDA), Middle East Technical University (METU)

ACADEMIC SERVICE

Program Committee Member #papers reviewed

- ECML/PKDD 2022, 2021 (Reviewer) `4 papers, 1 paper`
- NeurIPS 2022 (Volunteer)
- IEEE PIMRC 2019 (Reviewer) `2 papers`

PROJECTS

- Wikicapraim: a replica website of Wikipedia to collect clicks for prediction of "user navigation traces" `MediaWiki`
- HOBORODH: A run & dodge game with a racing car modeled after Lamborghini Aventador and designed to avoid molotov cocktails and bombs `C++` `OpenGL`
- Converting *Modular Accident Analysis Program* data to database query formats `C++`
- Predicted user ratings for films using Collaborative Filtering `Python`
- Personalized Daily Activity Recognition by analyzing data collected from accelerometer and gyroscope sensors using SVM & Logistic Regression `Matlab`
- Inference of Attributes from Crowd-sourced Annotations for Fake News Detection `Python`
- Diagnosis of cancer using Decision Tree Learning and Ada-boost ensemble classifier
  Text classification using K-Nearest Neighbor and Naive Bayes. `Python`
- Mega Structure Modeling - *Himeji Castle*, Lighting and Texture, Ray Tracing `Graphics`
- Collision Avoiding Robot using E-puck
- Implementation of Transport, Network and DLL Layers of OSI model `Java/C`
- A content management system (CMS) for college websites `PHP` `MySQL` `CodeIgniter`
- Implementation of thread synchronization, scheduling, multiprogramming, caching, process and virtual memory management of NACHOS `Java`
- Implementation of a Compiler (Symbol-Table, Lexical Analyzer, a Parser and an Intermediate Code Generator) for a subset of Pascal `Lex` `YACC` `C++`
- Four in a Row: A LAN multi-player game using Socket Programming `Java`
- Hangman: Word guessing game with enhanced audio & visual interface `C`

Last updated on February 2, 2023.