

KMA SOLAIMAN

CONTACT INFORMATION	Ph.D. Candidate Department of Computer Science Purdue University West Lafayette, IN 47907	 B132 #5 Lawson Computer Science Building  +1 765-775-8230  https://ksolaiman.github.io/  ksolaima@purdue.edu
RESEARCH INTERESTS	Multi-modal Information Retrieval, Heterogeneous Data Mining, Novelities in Machine Learning, and Data Management Systems.	
EDUCATION	Purdue University <i>Ph.D. in Computer Science</i> Advisor: Bharat K. Bhargava Purdue University <i>M.Sc. in Computer Science</i> Bangladesh University of Engineering and Technology (BUET) <i>B.Sc. in Computer Science and Engineering (CSE)</i> GPA: 3.79/4.00 (Ranked 16 th in a class of 153 students)	Fall 2016 - Spring 2023 Fall 2022 July 2014
TECHNICAL SKILLS	★ INDUSTRY KNOWLEDGE: Machine Learning, Data Science, Data Analysis, NLP, Computer Vision, Reinforcement Learning, Deep Learning, Graph Representation Learning ★ TOOLS AND TECHNOLOGIES: <ul style="list-style-type: none">• Software Stacks: Docker, Kafka, Conda, Git (Version Control), Amazon Web Services (AWS)• Programming Language: Python, Java, C, C++, 80x86 Assembly, Prolog, PHP, HTML, XML, CSS, JavaScript, Linux Shell Programming• Deep Learning Frameworks: PyTorch (<i>High</i>), Caffe, Matlab (<i>Moderate</i>)• Database Systems: Oracle, Microsoft SQL Server, MySQL, PL/SQL• Tools and Software: CSim, Matlab, UML, ERD, Weka, Ajax, JQuery, Lex, Yacc, OpenGL, L^AT_EX, Google API, Jupyter Notebook• Operating Systems: Linux, Unix, Windows	
COLLABORATIONS	Massachusetts Institute of Technology (MIT), University of Southern California (USC-ISI), Northrop Grumman Corporation (NGC), Institute for Defense Analyses (IDA)	
RESEARCH EXPERIENCE	Research in Applications for Learning Machines (REALM). <i>Bharat K. Bhargava, Michael Stonebraker</i> This project focuses on finding similar and anomalous data sources and patterns across heterogeneous data sources for delivering relevant information to user for real-time contextual awareness given current user information need. Collaborating with local authorities, we worked on finding a real-time scalable solution for <i>missing person search</i> where we worked with real-world noisy and high dimensional data collected by the agencies. <ul style="list-style-type: none">• Designed a <i>scalable cross-modal querying method</i> based on relational schema• Proposed two <i>weakly supervised methods</i> for <i>cross-modal information retrieval</i>• Proposed a novel human attribute recognition model from unstructured text and benchmarked attribute recognition models for video and image	

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

An agent in any learning-based systems faces many novel, unknown, or out-of-distribution inputs in real world environment. This work developed theories on novelty characterization, novelty detection, and novelty adaptation for learning-based systems with focus on reinforcement-learning based agents for playing multi-agent games (modeling real world collusion, co-operation, and adversary).

Understanding Political Bias in News Articles using Social Media. (Dan Goldwasser)

- Designed representations for newspaper articles using Contextualized Language Models
- Designed joint representations for text and social information
- Text classification for political bias with vanilla and neural network based classifiers
- Building weakly supervised models for collective classification using Probabilistic Soft Logic

Adversarial Attacks on Neural Networks. (Miguel Villarreal Vasquez) Spring 2019

Ran experiments to tackle the VGG-FACE deep neural network trojaning attack by introducing a healing dataset from Wild dataset and retraining it. Original attack pollutes the model by trojan trigger generation and retraining the model with reversed engineered training data.

Unsupervised Learning (Advisor: Md. Monirul Islam) Feb 2013 - Jul 2014

- Proposed a novel clustering algorithm for irregular and complex shaped data with only a single variable parameter, *filter-width* and described an empirical method to dynamically find optimal value of that parameter.
- Developed extension to Weka platform and showed performance comparison of our algorithm with K-means, EM, Cobweb & Classit, FarthestFirst, DBSCAN & Hierarchical Clustering algorithm across data sets like Aggregation, Compound, Jain, Spiral, and others.

Analysis & Visualization of Road Accident Data Oct 2011 - May 2013

- Implemented a novel web interface for collection of road accident data in Bangladesh and for performing dynamic data analysis of road accidents with contributing factors using Google APIs.

Data Mining, Complex Network Analysis

- Developed methods for finding hand-gesture patterns in smart watch sensor data using Recurrent Neural Networks, specifically LSTM. (Supervisor: He Wang) Sep 2017 - Oct 2017
- Investigated *TribeFlow* for mining and predicting user preferences using website hyperlink structure in Wikipedia. Jan 2017 - Jul 2017

PUBLICATIONS

- (Conference) KMA SOLAIMAN and B. Bhargava, Multi-modal Information Retrieval for Systems with Explicit Information Needs and Object Properties (Femmir), *Submitted in SIGMOD 2023*.
- (Journal) 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).
- (Symposium) K. SOLAIMAN AND B. BHARGAVA, Open-Learning Framework for Multi-modal Information Retrieval with Weakly Supervised Joint Embedding. *AAAI Spring Symposium*, March 2022. [Link]
- (Symposium) K. SOLAIMAN AND B. BHARGAVA, Measurement of Novelty Difficulty in Monopoly. *AAAI Spring Symposium*, March 2022. [Link]
- (Conference) A. NESEN, KMA SOLAIMAN AND B. BHARGAVA, Dataset Augmentation with Generated Novelty, *International Conference on Transdisciplinary AI (TransAI)*, IEEE, 2021.
- (Workshop) M. STONEBRAKER et al, Surveillance Video Querying With A Human-in-the-Loop, *Workshop on Human-In-the-Loop Data Analytics (HILDA) with SIGMOD*, 2020.
- (Workshop) 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*, co-located with VLDB, Part of the *Lecture Notes in Computer Science (volume 11721)* book series, Springer, 2019. **Co-first authors.

- (Conference) S ROY, K SOLAIMAN, C LI, D GOLDWASSER, Identifying Bias in News Narratives Using Distant Supervision, (**IWCS**), 2019. [Submitted Work]
- (Conference) KMA SOLAIMAN, MD. MUSTAFIZUR RAHMAN, AND NASHID SHAHRIAR, AVRA BANGLADESH: Collection, Analysis & Visualization of Road Accident Data in Bangladesh, *International Conference on Informatics, Electronics & Vision (ICIEV)*, IEEE, 2013.
- (Poster) KMA SOLAIMAN, AHMED AL MUZADDID, Minimal Parameter Clustering of Complex Shape Dataset with High Dimensional Dataset Compatibility, *Presented at First Undergraduate Thesis Poster Presentation, CSE, BUET, 2014*.
- (Undergraduate Dissertation) Minimal Parameter Clustering of Complex Shaped & Different Sized Dataset with Noise & Outlier Handling (MPCACS).

PROFESSIONAL **Research Assistant** Spring 2020 - Present
 EXPERIENCE *Purdue University, Supervisor: Bharat K. Bhargava*

Graduate Teaching Assistant

Computer Science, Purdue University

- * CS18000 : Problem Solving and Object Oriented Programming Fall 2016 - Fall 2017
- * CS25100 : Data Structures Spring 2017 - Fall 2018
- * CS54300 : Introduction To Simulation & Modeling Of Computer Systems Spring 2019
- * CS44800 : Introduction To Relational Database Systems Fall 2019
- * CS53600 : Data Communication and Computer Networks Fall 2022

Lecturer, *Computer Science & Engineering,*
Ahsanullah University of Science & Technology

Oct 2014 - Jul 2016

ACADEMIC **Program Committee Member / Reviewer** #papers reviewed
 SERVICE ** ECML/PKDD 2022* 4 papers
** ECML/PKDD 2021* 1 paper
** IEEE PIMRC 2019* 2 papers

COURSE **Wikicaprism: Replica of Wikipedia to collect “user navigation traces”.** MediaWiki
 PROJECTS **HOBORODH: A run & dodge game with a racing car modeled after Lamborghini Aventador and designed to avoid molotov cocktails and bombs.** 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 of OSI Layers Java/C
4 bit pipe-lined microprocessor, capable of executing 28 specific instructions Quartus
A content management system (CMS) for college websites PHP MySQL CodeIgniter
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
