KMA Solaiman

Computer Scientist, Purdue University

3689 Desoto Dr
West Lafayette, IN 47906
+1 (765) 775-8230
ksolaima@purdue.edu
ksolaiman.github.io/

My research intersects with **machine learning, multimodal information retrieval, open-world learning,** and **data management systems**. I built systems and techniques to enable scalable and robust multimodal data processing in open-world environment, while working with practical applications and end users. Combining insights from computer vision, nlp, representation learning, and information retrieval, I design techniques to enable practical cross-modal retrieval system via feature-centered knowledge accumulation. Additionally, my research focused on acquiring fundamental insights that facilitate dependable open-world learning, which can operate securely and flexibly in the presence of dynamic and unforeseen data streams. My work has appeared in top data management venues like IEEE, AAAI, SIGMOD, and VLDB. I have collaborated with multiple industry and academic partners such as MIT, NGC, USC, and DARPA.

Education

July 2023 **Ph.D. in Computer Science**, *Purdue University*, West Lafayette, IN.
- Advisor: Bharat K. Bhargava
- Mentor: Michael Stonebraker (MIT)
- Thesis: *Situational Knowledge on Demand in Open-world Environment*

Dec 2022 **M.Sc. in Computer Science**, *Purdue University*, West Lafayette, IN.
- Area: Machine Learning and Databases

July 2014 **B.Sc.**, *Bangladesh University of Engineering and Technology*, Dhaka, Bangladesh, GPA: 3.79.
- Thesis: *Minimal Parameter Clustering of Complex Shaped and Different Sized Dataset*
- Computer Science and Engineering
- Class Rank: 16/153

Professional Experience

2019-2022 **Purdue University**, *West Lafayette, IN, USA*, Research Assistant.
- Supervisors: Bharat Bhargava, Michael Stonebraker
- Projects: NGC REALM, DARPA SAIL-ON

2016 – 2019, 2022 – 2023 **Purdue University**, *West Lafayette, IN, USA*, Teaching Assistant.
* Graded 2+ projects and exams each semester for 90+ students with detailed feedbacks
* Instructed in labs and PSOs with ∼30 students (from freshmen to graduate levels)
* Designed homeworks, assignments and exams
* Mentored students for final course projects in OOP and reproducing papers in Computer Networks
* Course development for OOP and Simulation & Modeling

Aug 2014 – Jul 2016 **Ahsanullah University of Science and Technology**, *Dhaka, Bangladesh*, Full-time Lecturer.
* Conducting labs and supervising group projects in Database, Networking and Software Engineering
* Student advising, participating in accreditation, and curriculum development for courses in CS
* Handled 12-18 credit hours each semester as primary instructor with academic services

2014 **United International University**, *Dhaka, Bangladesh*, Full-time Lecturer.

Research Experience

2019-2021 **Research in Applications for Learning Machines (REALM)**, *Purdue, MIT and NGC.*
Co-advised by: *Michael Stonebraker*
- In collaboration with local police, a *scalable cross-modal querying method* was built using relational dbms and data fusion. The prototype was presented for finding a real-time scalable solution for *missing person search* with real-world noisy and high dimensional dataset. [J1]

- Proposed a *cross-modal matching* method based on graph matching and co-ordinated representation learning. [C4]
- Proposed a *weakly supervised method* for *multimodal information retrieval* based on contrastive learning and representation learning. [W4]
- Proposed a novel *human attribute recognition model from unstructured text* using Word2Vec, SBERT and WordNet. [C4]
- Scraped ∼10K tweets originated from Cambridge, MA and used similarity search (e.g., LSI, LDA) to identify high level objects in tweets similar to traffic videos. [W1]

2021-2022 **Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON)**, *USC-ISI*.

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

2018 **Understanding Political Bias in News Articles using Social Media**.
Supervised by: *Dan Goldwasser*

- Experimented with different language models for representating news articles and built text classifiers for political bias.
- Designed a joint representation learning method for identifying political stance in newspaper articles using weak supervision from tweets.
- Used Amazon MTurk to design a bias detection task for annotating newspaper articles with political bias.

Oct 2019 **Adversarial Attacks on Neural Networks**, *Purdue University*.
Collaborators: *Miguel Villarreal Vasquez*

- Experimented 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.

2013 - 2014 **Unsupervised Learning for Complex Datasets**, *BUET*.
Advisor: *Md. Monirul Islam*

- Proposed a *novel clustering algorithm for irregular* and *complex shaped dataset* with a single parameter, *filter-width*. [P1]
- 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 other clustering algorithms: K-means, EM, etc.

2011 - 2013 **Analysis and Visualization of Road Accident Data**, *BUET*.
Collaborators: *Md. Mustafizur Rahman* and *Nashid Shahriar*

- Implemented a novel web interface for collection of road accident data in Bangladesh. [C1]
- Performed *data analysis of road accidents* to compare and identify prime contributors for accidents e.g., rural vs urban using Google APIs. [C1]

2017 **Data Mining and Complex Network Analysis**, *Purdue University*.

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

Publications

Peer-reviewed Conference (C), Journal (J), Workshop and Symposium (W) Papers

(** indicates **co-first** authors)

IEEE 2022 [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*.

AAAI 2022 [W4] **K. Solaiman** and B. Bhargava, *Open-Learning Framework for Multi-modal Information Retrieval with Weakly Supervised Joint Embedding*, **AAAI Spring Symposium** on Designing Artificial Intelligence for Open Worlds, *March 2022*.

AAAI 2022 [W3] **K. Solaiman** and B. Bhargava, *Measurement of Novelty Difficulty in Monopoly*, **AAAI Spring Symposium** on Designing Artificial Intelligence for Open Worlds, *March 2022*.

IEEE 2021 [C3] A. Nesen, **K. Solaiman** and B. Bhargava, *Dataset Augmentation with Generated Novelties*, IEEE **TransAI**, *2021*.

SIGMOD'20 [W2] Michael Stonebraker, Bharat Bhargava, Michael Cafarella, Zachary Collins, et al., *Surveillance Video Querying With A Human-in-the-Loop*, Workshop on Human-In-the-Loop Data Analytics (**HILDA**) with **SIGMOD**, *2020*.

VLDB 2019 [W1] S. Palacios, **K. Solaiman****, P. Angin, A. Nesen, B. Bhargava, Z. Collins, A. Sipser, M. Stonebraker, *SKOD: A Framework for Situational Knowledge on Demand*, **POLY** at **VLDB**, *Springer 2019*.

IEEE 2013 [C1] **Kma Solaiman**, MM Rahman, and N Shahriar, *AVRA BANGLADESH: Collection, Analysis & Visualization of Road Accident Data in Bangladesh*, IEEE International Conference on Informatics, Electronics & Vision **(ICIEV)**, *2013*.

Conference (C) and Journal (J) Preprint/ Submissions and Posters (PS)

AI 2023 [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*.

SIGMOD'23 [C4] **K. Solaiman** and B. Bhargava, *Multi-modal Information Retrieval for Systems with Explicit Information Needs and Object Properties (FemmIR)*, **SIGMOD** *2023*.

IWCS 2019 [C2] S Roy, **K Solaiman**, C Li, D Goldwasser, *Identifying Bias in News Narratives Using Distant Supervision*, International Conference on Computational Semantics (**IWCS**), 2019.

BUET 2014 [PS1] **Kma Solaiman** and AA Muzaddid, *Minimal Parameter Clustering of Complex Shape Dataset for High Dimensional Dataset*, **BUET CSE Thesis Poster Presentation**, *2014*.

Teaching Experience

2016 – 2023 **Purdue University**, *Teaching Assistant*.

- CS 180: Problem Solving and Object-Oriented Programming (OOP) — *3 semesters*
- CS 251: Data Structures — *3 semesters*
- CS 448: Introduction to Relational Database Systems — *2 semesters*
- CS 543: Simulation & Modeling of Computer Systems — *Graduate Course, Spring 2019*
- CS 536: Data Communication and Computer Networks — *Graduate Course, Fall 2022*

Fall 2019 **Purdue University**, *Guest Lecturer*.

- CS 590: Situational Awareness, Adversarial ML, and Explainable AI — *Graduate Course*
- CS 547: Information Retrieval — *Graduate Course*

2014 – 2016 **Ahsanullah University of Science & Technology, United International University**.

- Primary Instructor for the undergraduate course, Programming Language — *106 students*
- Primary Instructor for the undergraduate course, Network Programming — *143 students*

- Primary Instructor for the undergraduate course, Database *132 students*
- Primary Instructor for the undergraduate course, Simulation and Modeling *50 students*
- Primary Instructor for the undergraduate course, Graphics *50 students*

Research Mentoring

- Kevin Kochpatcharin (Purdue BSc+MSc → Five9)
- Tomáš Hrdlovics (Purdue MSc → WePay)
- Harshit Singh (Purdue BSc → Cisco)
- Sharuna Anandraj (Purdue MSc → Meta)
- Varsha Venkata Krishnan (Purdue MSc)
- Myeongsu Kim (Purdue MSc → Purdue PhD)
- Rumela Ghosh (Purdue MSc → Amazon)
- Rabia varol (METU → TUM MSc)
- Merve Yaman (METU BSc)
- Doruk Gercel (METU BSc → TUM MSc)
- Aaron Sipser (MIT BSc → Software Engineer)
- Zachary Collins (MIT BSc → Facebook)

Key Collaborators

MIT	**Massachusetts Institute of Technology**	Michael Stonebraker, Tao Sun
UMichigan	**University of Michigan**	Michael Cafarella
USC-ISI	**University of Southern California, Information Sciences Institute**	Mayank Kejriwal
UT Dallas	**University of Texas at Dallas**	Eric Kildebeck
IDA	**Institute for Defense Analyses**	Josh Alspector
NGC	**Northrop Grumman Corporation**	Jim MacDonald, Jason Kobes
Purdue	**Purdue University**	Alina Nesen, Xavier Tricoche, Shafkat Islam, Ganapathy Mani, Miguel Villarreal Vasquez, Servio Palacios, Vaneet Aggarwal
METU	**Middle East Technical University**	Pelin Angin
UBD	**Universiti Brunei Darussalam**	Sandhya Aneja, Nagender Aneja
	Visiting Scholars	Ruy Oliveira, Arun Kumar

Technical Skills

Relevant Coursework	Crowd-sourcing and Social Computing, AI meets Sustainability, Machine Learning, Data Mining, Distributed Database Systems, Algorithms
Fluency	Python, Java, C++, PyTorch, PostgreSQL, MySQL, LaTeX, Jupyter Notebook

Awards and Honors

2018	Graduate School Summer Research Grant, Purdue University
2009-2012	University Merit Scholarship, Bangladesh University of Engineering and Technology
2008-2011	University Stipend, Bangladesh University of Engineering and Technology
2009-2011	Dean's List Award, Bangladesh University of Engineering and Technology

Grants

2023-2026 **DARPA Triage Challenge (DTC) with a budget of $1.5M/year**, Submitted
Autonomous Triage Agent (for secondary triage) to identify physiological features of life-threatening injuries in mass casualty incidents.
Principal Investigators: Xavier Tricoche, Bharat Bhargava, Tianyi Zhang, and Eric Kildebeck

2022-2026 **DARPA In the Moment BAA, $5M**, Abstract accepted
Hippocrates: Human-aligned Autonomous Triage System.
Principal Investigators: V. N. Venkatakrishnan, Xavier Tricoche, Ardhendu Tripathy and Daniel Shapiro

2023-2026 **Submitted to NSF, $599,000**
VecDB: A Cloud-Native Vector Data Management System.
Principal Investigators: Jianguo Wang and Bharat Bhargava

2022-2023 **Submitted to DARPA (Joint with USC-ISI), $300K**
Stabilizing Hostilities through Arbitration and Diplomatic Engagement (SHADE).

2022-2024 **Sandia National Lab STARCS**
Computing Reconfiguration for Resilient Space Platforms.
Principal Investigators: Chris Jenkins and Bharat Bhargava

2019-2022 **Northrop-Grumman Corporation (NGC) REALM Consortium, $460K/year with MIT, CMU, Stanford**, Purdue awarded $200K for last 3 years, Ranked highest in nationwide competition
Advances in Learning Machines from Sensing to Acting for Mission Objectives.
Principal Investigators: Bharat Bhargava, Michael Stonebraker, Aarti Singh, and Matei Zaharia

2019-2023 **DARPA SAIL-ON (joint with USC-ISI), $1,245,990 (Purdue shared over 42 months)**
Generating Novelty in Open-world Multi-agent Environments (GNOME).
Principal Investigators: Mayank Kejriwal, Bharat Bhargava, and James MacDonald

Open Source Software and Demo

2022 **Find-Them**
Video demonstration of the system prototype for 'Applying Machine Learning and Data Fusion to the Missing Person Problem' (IEEE Computer 2022).
https://tinyurl.com/find-them

2020 **Surveillance Video Querying Engine (SurvQ)**
Contains the repository for the video querying engine (SIGMOD 2020). The artifact describes the querying and results UI, the video feature extractor, the video processor module, the relational dbms query processing, and the code for ingestion to delivery workflow.
https://github.com/skod-ng/ and SurvQ-Demo

2019 **Situational Knowledge on Demand (SKOD)**
Contains the repository for the SKOD framework (VLDB 2019). Video and Tweets ingestion process are implemented in twitter-kafka-docker and videos-docker. The knowledge graph along with the software frontend can be found in knowledge-graph.
https://github.com/purdue-gask and SKOD-Demo

Public Media

May 2019 **The Right Information at the Right Time**
Purdue CS News.
https://www.cs.purdue.edu/news/articles/2019/bhargava-realm-ng.html

Keynote Talks

May 2022 **Detect, Characterize, and Accommodate Novelties in AI systems**
International Semantics Intelligence Conference (ISIC).
Presented with Bharat Bhargava

Feb 2021 **REALM: Situation Knowledge on Demand (SKOD)**
International Semantics Intelligence Conference (ISIC).
Presented with Bharat Bhargava

Invited Talks

July 2021 **Adaptable AI Systems to deal with Novelties and Attacks**
Artificial Intelligence Campaign Tech Talk.
Presented with Bharat Bhargava

May 2021 **Information-Theoretic approach for determining difficulty of adaptation to novelty in Monopoly**
Novelty Working Group for SAIL-ON.
Hosts: Joshua Alspector and Pat Langley

Feb 2021 **Situation Knowledge on Demand (SKOD)**
Cyber Defense Engineering and Research group, JPL-NASA.
Hosts: Arun Viswanathan and Jeremy Pecharich

Dec 2020 **SKOD Research Progress and Future of Multimodal Information Retrieval**, Extracting relations between features, objects and entities; multimodal data association; logical understanding of similar features; and automated context knowledge generation of multimodal data
NGC REALM Year-End Meeting.
Host: Reid Hyland

Aug 2019 **SKOD: A Real-time Urban Information System**
Northrop Grumman TechFest.
Host: Keyla Contreras-Cottin

Professional Service

Program Committee Member

2022, 2021 External Reviewer for European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'22, ECML PKDD'21)

2022 Volunteer for the Conference on Neural Information Processing Systems (NeurIPS'22)

2019 External Reviewer for IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'19)

Workshop Organization

2021 **Student organizer**, Workshop on Novelties in Open World, in conjunction with International Semantic Intelligence Conference (ISIC-2021)

2020-21 **Active participiant** in the biweekly meetings of **DARPA SAIL-ON NWG** (Novelty Working Group) for 2 years

2014 **Lead organizer**, Seminars on *Higher Education* and *Careers in the Industry* for undergraduate students at BUET

Outreach Activities

2016 Participated in the **accreditation of undergraduate studies** at AUST

2016 **Undergraduate Thesis/ Project Committee Member**, Ahsanullah University of Science and Technology, Computer Science