Research Statement *KMA Solaiman*

Real-world use-cases in data-centric applications (including societal, healthcare, or education) with minimal computational resources often have an unprecedented influx of unstructured and noisy data from multiple sources and modalities. Extraction of meaningful information from such heterogenous and changing datasets requires achieving the complementary functionalities of *cross-modal matching, scalable data management system (mostly search engines and databases),* and *situationally-aware data recommendation.* My research goal is to understand how these functionalities can be achieved by designing interactive algorithms and robust systems that solve the problem of **situational knowledge on demand in open-world** from multiple forms of information, regardless of whether presented as text, images, videos, audio, or other modalities, while achieving data-driven and resource-aware *data management* and *data integration* capabilities.

My later works branched into developing scientific principles to *quantify* and *characterize novelty* (significant and unexpected events) *in open-world domains*, while creating scalable and efficient AI systems that *react to novelty* in those domains.

Developing intelligent AI systems including situational knowledge recommenders for open-world environments involves tackling multiple *system design challenges*:
(1) Resource-aware Data Management: Data management systems require a comprehensive understanding of the data properties, user requirements, and limitations imposed by open-world to achieve optimal performance and scalability for multimodal data recommendation. Emerging multimodal applications impose challenges for traditional system (hardware and software) design choices ranging all the way from input (heterogeneous sources, context and modalities) and output (delivery-on-time, quick throughput and diverse users), changing information needs (knowledge base creation and query), to computational resources (lack of annotations, domain-specific feature extractors or human resources). For instance, we showed that transfer learning for fine-grained semantic concept extraction from videos turned out to be ill-suited in large-scale systems [4].
(2) Data Integration: Data integration from various sources to answer queries over a single view of the data to users, is confronted with a multitude of heterogeneity issues. These problems arise from differences in data attributes names that hold similar data [9, 7], communication problems, and variations in data schema and types [6]. As data volume increases and the necessity to share existing data intensifies, data integration becomes more prevalent. My first approach for data integration, EARS, delivers integrated query results over time using a mediation approach and schema mapping [9], solving the problem of scalability and quick throughput. The second approach, FemmIR, learns a co-ordinated graph representation of the data samples comprised of their semantic features to deliver approximate matches [7]. The third approach, WesJeM, uses Contrastive Learning to embed data-objects and their semantic properties in a high-dimensional space using higher-level semantic features in a data sample as weak labels [6], allowing zero-shot similarity matching and data discovery of multimodal data in open-world environments.
(3) Dealing with Open-world Novelties: To construct intelligent AI systems, it is necessary to adapt to evolving scenarios. However, conventional AI systems face restrictions when it comes to managing unexpected events or "novelties" that were not previously seen or modeled. We need to characterize, detect and adapt to novelties at various stages of the AI life cycle, such as data integration, relevance learning, environment modeling, feature extraction, and inherent domain properties. Novelty characterization and difficulty estimation is required in a plethora of AI systems ranging from multimodal information retrieval [6] (distribution change and concept drift), dataset complexity [1], to visual (object detection in video and image [2]) and planning domains (games [5] and war).

My work has shown positive results on open societal problems that previously required large-scale human endeavor and computational resources, such as Missing Person Search [9], Dataset Complexity Estimation [1], Search and Rescue in Disasters [8], and Medical Triage. The impacts of my work in academia span across diverse areas, including Multimodal Information Retrieval, Data Discovery, Building Robust AI Agents, and Data Completion, thereby creating a significant impact.

My research on Novelty Adaptation and Situational Knowledge Delivery has the capability to **transform information retrieval for everyone by empowering AI systems** to promptly and precisely retrieve the information required by users, even in situations where the resources are minimal, data sources are in a state of constant change, and involve various modalities, interconnections, and predictive arguments.

RESEARCH PHILOSOPHY. I believe that the ultimate objective of AI is not just to enhance performance on isolated tasks but to complement human abilities in solving persistent issues in real world, minimize human labor through responsible action, harness the power of data for our benefit, and make decisions that have long-lasting positive effects. My work is driven by a desire to integrate these ideas into a context that prioritizes social good and encourages the consumption and sharing of healthier information. This involves leveraging multimodal techniques to better understand complex situations, adapting to different situations, and providing trustworthy and easily understandable information for humans.

Resource Aware Data Management for Multimodal Applications

In designing data management systems for modern applications, such as missing person search, disaster resource management, triage, and emotion recognition, the focus has shifted to account for data-at-rest and streaming input while ensuring scalability to handle increasing information needs and data ingestion. To address these challenges, during our collaboration with local police department and MIT for building Missing-person Query engine [9] and Human-in-the-Loop Video Querying system [4], we proposed a novel multimodal knowledge querying system called **SKOD** (Situational Knowledge on Demand) [3, 4]. SKOD leverages entity-centric higher-level semantic concepts (such as objects, object types, physical relations, e.g., a person wearing blue shirt, time and place of an incident), and the functionalities of distributed systems and RDBMS to query domain-specific information needs in practical multimodal applications. SKOD was developed in collaboration with Northrop Grumman, MIT, and CMU and demonstrated at Northrop Grumman TechFest in 2019. The project has been funded by Northrop Grumman for three consecutive years since 2019, renewing the funding every year.

Heterogenous Data Ingestion, Scalability, and Delivery-on-demand

We used Postgres as the backend architecture for both data storage and on-time delivery. Building on top of the RDBMS allows us to scale to practical data volumes, as well as using the querying interface with query-by-example and query-by-features dramatically lowers the human costs of multimodal and visual domain search [4]. For consuming data from heterogeneous sources (both at rest and streaming), I integrated Kafka producers and consumers on top of SKOD [3]. Any query to the system was formulated and considered as an *incident in real life*. For delivery-on-demand from incomplete modalities, we used Postgres Trigger functionality, which is activated whenever an insert occurs that matches a certain incident (any matching data). This feature allows us to deliver incomplete information need and complete it later when new matching data is encountered, while being capable of adapting to changing information requirements. Queries in SKOD can be both standing queries or one-shot queries. To deal with the changing requirements, we proposed to build a ***query-drive knowledge base*** for each user, where all queries can relate to a single incident. SKOD speeds up the data delivery by storing frequent incidents by caching hot queries, and recently used data.

Resource-constrained Feature Extraction

Task-specific querying systems face challenges during the data preparation stage due to low-quality data sets and a lack of labeled training samples. Although large-scale language models have made significant progress, there exists very few task-specific attribute extractors for text. Our team addressed these issues in SKOD by implementing a *priority polling system* that selects candidate data samples for feature extraction from videos and images, instead of immediately processing features for batch inputs. This feature, coupled with trigger functionality, enables us to provide information needs on-demand and complete them with time. Additionally, we developed a **cloth-color extractor** for videos using common-sense reasoning and color and shape analysis [4] on top of YOLO. To identify attributes in unstructured text, I propose a model called **HART** [7], which solves the problem in two stages: (i) **candidate sentence identification** by transforming the problem into a similarity-search problem using pre-trained language representation models (SBERT) and lexical knowledge bases, and (ii) **semantic attribute understanding** using syntactic characteristics and lexical meanings of the tokens in the candidate sentences. This approach can be generalized for any domain and lays the groundwork for intelligent document processing.

Label-efficient Data Integration at Higher Semantic Level

View-based Data Integration

Traditional data integration approaches suffer because of heterogeneity among data sources and incomplete modalities. Machine learning models for multimodal data fusion learn joint representations to exploit complementarity and redundancy of multiple modalities, but overlooks the information needs based on higher-level semantic concepts. With the use of Postgres trigger and by using a mediated schema for each queried incident, SKOD delivers integrated query results over time. Since the number of properties-of-interest are quite moderate, using similar approach to the *Global as View* data integration, I proposed to employ **schema mapping** between the *mediated schema* and *local schema* from different data sources. The proposed approach, **EARS** [9] adopts an entity-relationship-attribute schema for each new data source, and a wrapper is designed to translate the source schemas to the mediated schema. The queries are translated into conjunctive queries between features among data sources and a SQL-Join is performed at run-time to integrate all the relevant sources. Using the versatility of Postgres, we achieve the scalability and speed that is required for time sensitive use-cases, with minimal amount of computational resources.

Approximate Matching using Graph Representation Learning

While the SQL-JOIN based relational DBMS approach allows a lot of flexibility, it does not utilize the historical knowledge of previous queries, and cannot perform approximate matching. Considering the sensitivity of some open-world application domains, it is desirable to search for approximate relevance between different modalities and sources. Motivated by representation-invariant properties of graph representation models combined with the existing works on approximate graph matching techniques, I propose **co-ordinated graph representation learning of the data samples comprised of their semantic features** [7], where it learns to approximate a novel Edit distance metric, *CED*, based on the multiplicative comparison of the *Hierarchical Attributed Relational Graph* representations.

Weakly Supervised Metric Learning for Cross-Modal Matching

For real-world systems, *data discovery from heterogenous modalities* and *explanation of the relevant properties among similar data objects* is of equal importance. Since in these applications, manual annotation is not feasible or they lack annotation resources, we need alternative supervision techniques for cross-modal matching. Motivated by the advancement in translation and captioning models (video/audio $\rightarrow$ text), I propose to embed data-objects and their semantic properties in a high dimensional embedding space via Contrastive Learning. After extracting the interaction among entity-centric higher-level semantic features (such as, topics, events, entities, triplets) from texts and other translated modalities, a *data information network* is built by connecting data samples to their features via their interactions. Finally, I construct a structure-infused representation for the data-objects from all modalities in **WesJeM** [6], by jointly embedding the data samples, the features, and the available similarity labels, in a single space. For learning, I defined a multi-task learning objective capturing the interaction information, by aligning the representation of the data samples, defined by their textual content, with the representation of features, based on their common relations. For open-world environment where data and information-need keep changing, along with the dynamic data sources, WesJeM opens up the path for **Zero-Shot similarity matching** and **Data Discovery** of multimodal data.

ADAPTION TO OPEN-WORLD NOVELTIES

AI systems are often limited by their inability to handle unexpected events that are not part of their training data or well-defined environments. These significant changes or events are referred to as **'novelties'** under DARPA SAIL-ON project, and their characterization and adaptation is critical for real-world applications. To build robust and intelligent AI systems, I developed novelty characterization and adaptation techniques at various stages, including data integration and relevance learning, environment modeling, feature extraction, training, and domain or data level.

Novelty Characterization, Detection, and Difficulty Estimation

I characterized the **novelties encountered in multimodal information retrieval** in [6] and proposed how WesJeM can be adapted for changing data patterns and incomplete or noisy modalities in data integration and relevance learning stage. Moreover, motivated by the information-theoretic approach for difficulty estimation of novelties, I proposed an empirical framework for novelty characterization and difficulty estimation in **planning domains** [5]. For a reinforcement-learning based Monopoly agent, graphically modeling the environment to augment the state and action space allow to integrate graph edit distance as a novelty difficulty metric.

Robust Feature Extraction with Dataset Augmentation

The efficiency of entity-centric machine learning models in response to novelties depends on the efforts during the model training, design and data collection stages. We proposed a **novelty generation framework** [2] at the data preparation stage of training a model to assure its robustness and reduce the bias. We augmented the original dataset in a domain-agnostic and budget efficient manner with generated novelties for visual modalities, and improved the **novel object detection** performance with the augmented dataset.

Intrinsic Domain Complexity Estimation for Distributed AI Systems

Understanding of the inherent characteristics of the domain is essential for novelty characterization and model adaptability. We proposed an **application-independent domain complexity** measure for the AI systems in perception domain [1] using **federated learning as the reference paradigm** to handle distributed dataset operations. Building upon intrinsic dataset properties such as dimensionality, heterogeneity and sparsity for singular environment, we created a complexity metric *for the distributed environment*, showing efficacy for classification task.

FUTURE RESEARCH AGENDA

Currently, we are experiencing a thrilling era for multimodal information processing and robust AI research since it is highly supported by the core programs in NSF's Division of Information and Intelligent Systems (IIS) and by "Harvesting the Data Revolution (HDR^2)" idea - second wave of one of the 10 big ideas by NSF for long-term research.

My **long-term goal** is to create intelligent systems that can reason, learn and cooperate with humans to improve the standard of living by utilizing the vast amounts of data available in the modern era. My focus is to devise new algorithms and methods that can make a significant impact on society, leverage existing scientific advancements, and address real-world challenges. To that end, I plan to continue my research on *multimodal data management in real world* by approaching from the following directions:

User Preference Modeling

To complete the life-cycle of *situational knowledge delivery*, we still have challenges in modeling user's information need in a robust and efficient manner in multiple directions [9]: (1) user requirement is not always obvious or explicitly stated, (2) user can be interested in multiple types of events and knowledge bases with varying probabilities, (3) learning algorithms need to *adapt to changing user preferences with time*. I aim to develop novel algorithms using techniques such as active learning and reinforcement learning that can accurately capture and predict users' preferences based on their behavior, interactions, and feedback. Understanding the features that drive user preferences, and leveraging this knowledge to improve personalized recommendations and user experience, has applications in education (student advising, classroom teaching), e-commerce, healthcare, etc. To achieve this research goal, collaborations with researchers in **human-computer interaction, psychology, and marketing** will be essential.

Explainability and Trustworthiness in Data Recommendation

As the amount of multimodal data generated and consumed by users is increasing, there is a growing need for users to understand the basis for recommendations [7] and the saliency and trustworthiness of the information being consumed. This is especially important in sensitive domains such as *healthcare, finance, and legal decision-making* to allow for tracking, cross-checking with social contexts and verification. To achieve this goal, collaboration across multiple areas is necessary, including **data science, natural language processing, computer vision, human-computer interaction, and ethics**. With this, we can ensure that these models are designed with the user in mind, taking into account their cognitive and perceptual abilities. This collaboration can also lead to the *development of ethical guidelines and principles for designing trustworthy systems*, ensuring that users' rights and privacy are protected.

Privacy preserving Data Dissemination and Federated Learning

To address the growing concern over data privacy, particularly in medical and identity contexts, research in privacy-preserving multimodal data dissemination and federated learning is crucial, as identified in SKOD framework [3]. Further research to integrate the use of local data processing and remote federation with multimodal machine learning techniques is needed to ensure this new requirement in information processing, while understanding and formalizing the resource requirements. Collaboration across various fields such as **information security, statistics, data management, law, ethics, and public policy** is vital to advance research in this area.

Information Completion and Data Democratization

As data becomes increasingly important in all domains, there is a need for new techniques that enable individuals and organizations to efficiently extract insights from data and complete missing information. To address this challenge, future research should focus on developing advanced machine learning models that are able to perform well even with incomplete data, as well as methods for effective data integration and knowledge transfer within organizations. Collaboration is needed between **machine learning experts, data management specialists, and domain experts in various fields** to achieve a comprehensive and effective solution for data democratization and information completion.

COLLABORATION AND FUNDING

My future research vision requires collaboration with expert researchers in many fields, including natural language processing, computer vision, machine learning, data mining, social science, human computer interaction, systems and databases. I gained extensive expertise in overseeing and directing major projects, encompassing teams of over

12 individuals and collaborating with various universities and institutions. I led multiple masters and undergraduate students, collaborated with multiple Ph.D. students and coordinated with 5 professors from different universities to participate in the REALM project. I am fortunate to have close collaborations with professors from multiple universities and research institutes, such as Massachusetts Institute of Technology (MIT), University of Michigan (UMichigan), University of Southern California (USC), Information Sciences Institute (ISI), Institute for Defense Analyses (IDA), University of Massachusetts (UMass), Middle East Technical University (METU), etc. I also have had the fortune to work closely with researchers from databases and applications, along with end-users and program managers to conduct interdisciplinary research. I intend to maintain my current collaborations while actively cultivating new partnerships to advance the establishment of robust principles that underpin research in multimodal knowledge and novelty in learning models.

During my Ph.D., my work has been mainly supported by the Northrop Grumman Corporation (NGC), DARPA, ARFL, and Sandia National Lab. Additionally, I have contributed significantly to the writing of grant proposals, including idea generation, method design, idea illustration and visual aid creation, such as DARPA ITM project, and DARPA Triage Challenge. As a future faculty, I will continue to seek funding opportunities in the future from early career fellowships and various funding agencies (e.g., DARPA, ARL, AFRL, IARPA, NSF, NIH, DOE, DOD) and industries (e.g., NGC, Microsoft, IBM, Ford, Meta, Google, Intel).

REFERENCES

[1] Shafkat Islam*, **KMA Solaiman***, Ruy De Oliveira, and Bharat Bhargava. Domain complexity estimation for distributed ai systems in open-world perception domain. *Submitted to Artificial Intelligence*, 2023. ***Co-first Author**.

[2] Alina Nesen, **KMA Solaiman**, and Bharat Bhargava. Dataset augmentation with generated novelties. In *2021 Third International Conference on Transdisciplinary AI (TransAI)*, pages 41–44. IEEE, 2021.

[3] Servio Palacios*, **KMA Solaiman***, Pelin Angin, Alina Nesen, Bharat Bhargava, Zachary Collins, Aaron Sipser, Michael Stonebraker, and James Macdonald. SKOD: A framework for situational knowledge on demand. In *Heterogeneous Data Management, Polystores, and Analytics for Healthcare with VLDB*, pages 154–166. Springer, 2019. ***Co-first Author**.

[4] Michael Stonebraker, Bharat Bhargava, Michael Cafarella, Zachary Collins, Jenna McClellan, Aaron Sipser, Tao Sun, Alina Nesen, **KMA Solaiman**, and Ganapathy Mani. Surveillance video querying with a human-in-the-loop. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics with SIGMOD*, 2020.

[5] **KMA Solaiman** and Bharat Bhargava. Measurement of novelty difficulty in monopoly. In *Proceedings of the AAAI Spring Symposium on Designing Artificial Intelligence for Open Worlds*, 2022.

[6] **KMA Solaiman** and Bharat Bhargava. Open-learning framework for multi-modal information retrieval with weakly supervised joint embedding. In *Proceedings of the AAAI Spring Symposium on Designing Artificial Intelligence for Open Worlds*, 2022.

[7] **KMA Solaiman** and Bharat Bhargava. Multi-modal information retrieval for systems with explicit information needs and object properties (FemmIR). *Submitted*, 2023.

[8] **KMA Solaiman**, Md Mustafizur Rahman, and Nashid Shahriar. AVRA bangladesh collection, analysis & visualization of road accident data in Bangladesh. In *International Conference on Informatics, Electronics and Vision (ICIEV)*, pages 1–6. IEEE, 2013.

[9] **KMA Solaiman**, Tao Sun, Alina Nesen, Bharat Bhargava, and Michael Stonebraker. Applying machine learning and data fusion to the *Missing Person* problem. *IEEE Computer*, 55(6), 2021.