Multimodal Information Retrieval in Open-world Environment Right Information at the Right Time

KMA Solaiman

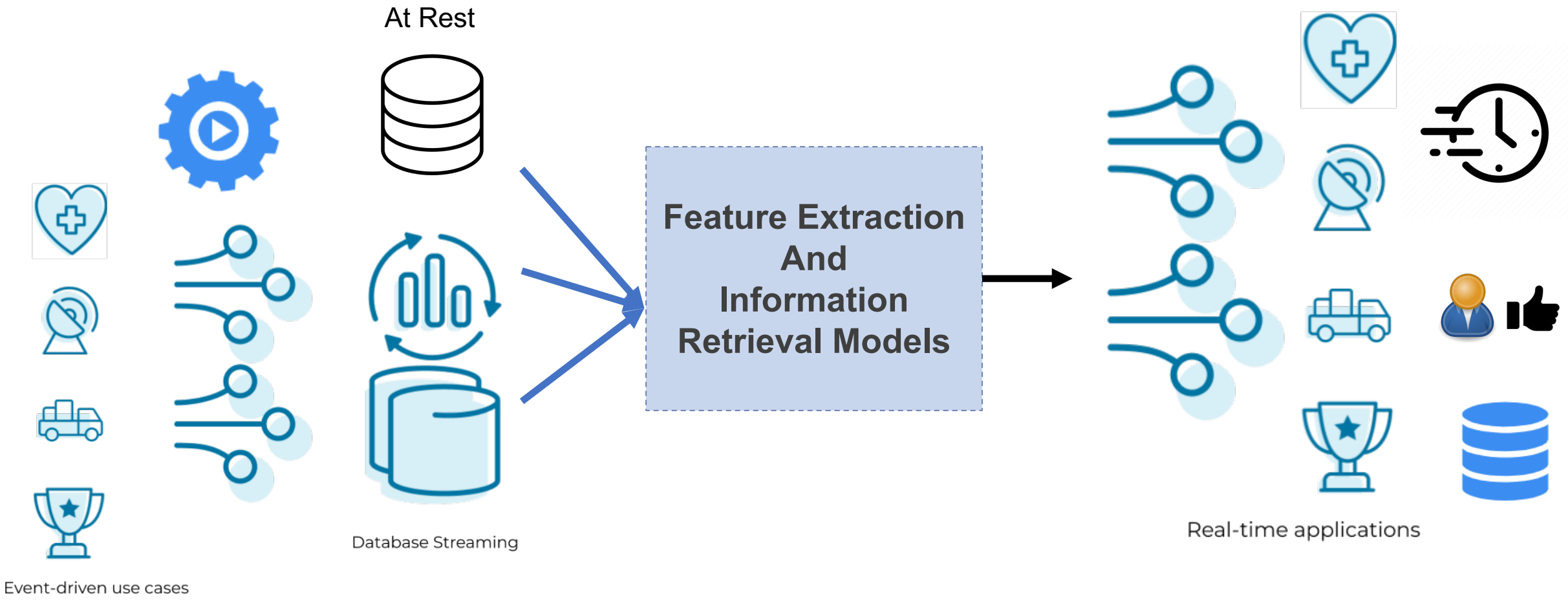
Purdue University

ksolaima@purdue.edu

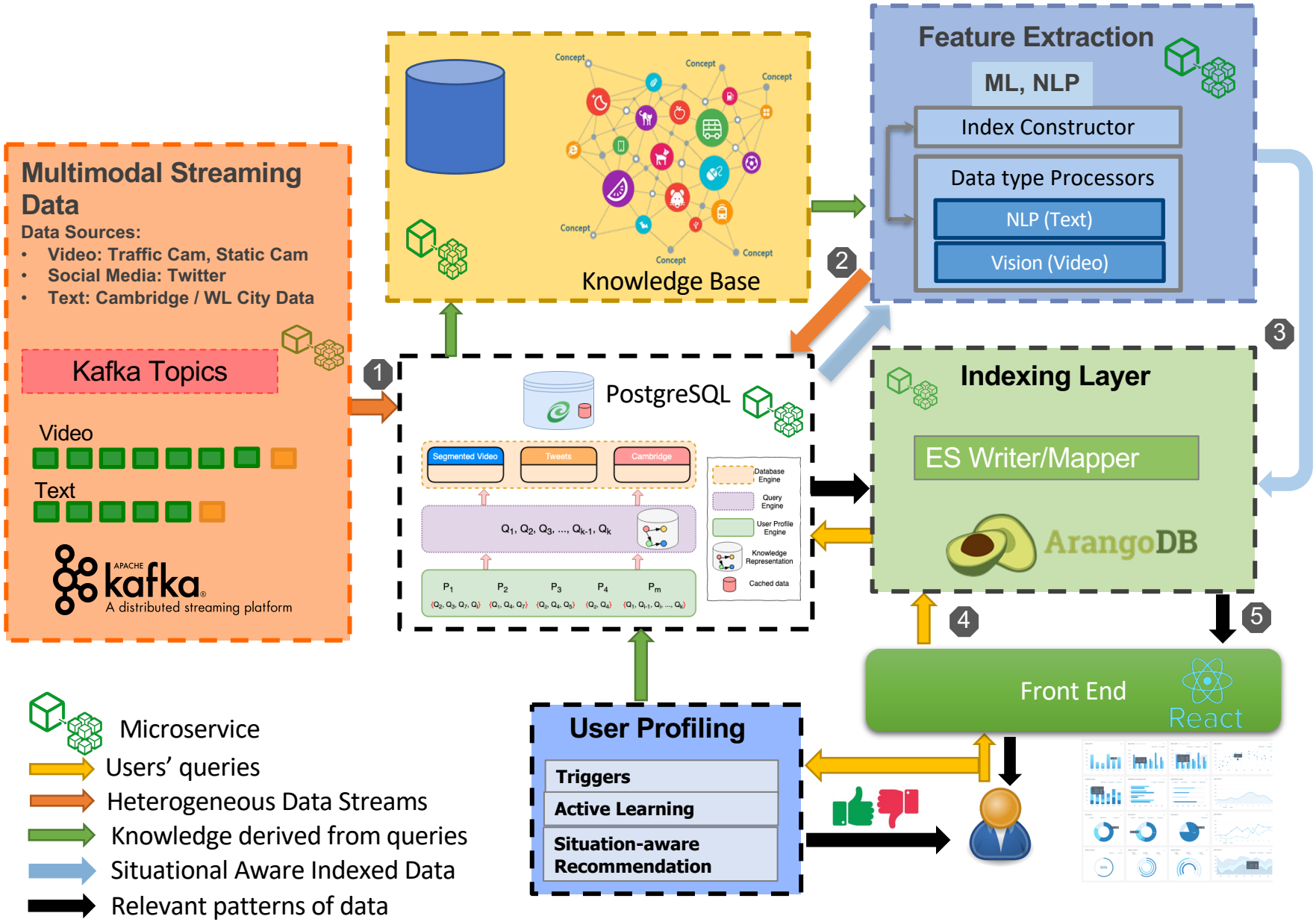
Problem formulation

- Build systems capable of satisfying multimodal information needs in open-world environments
- Make sense of data from novel and heterogeneous sources
- User satisfaction for multimodal information extraction in real-world depends on few pivotal aspects:
  - Consume and process streaming and at-rest heterogeneous data
  - On-demand data delivery
  - Handle lack of class labels
  - Resource constrained data preprocessing
  - Adapt to novel situations including changes in data and sources

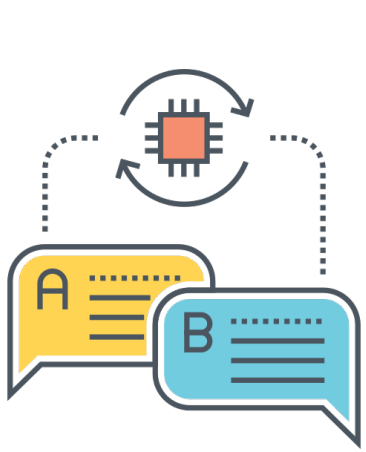
Challenge 1: Ingestion and Delivery

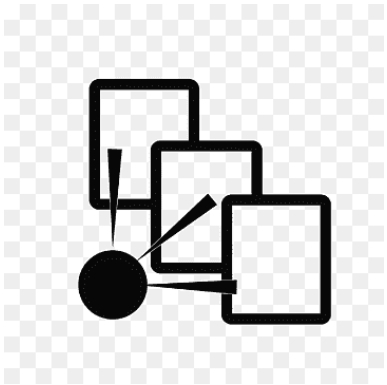


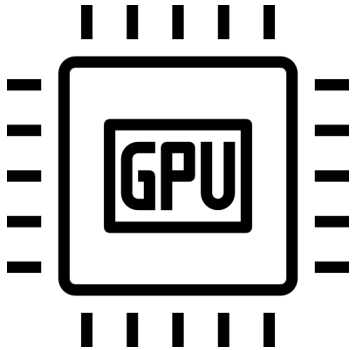
Situational Knowledge on Demand

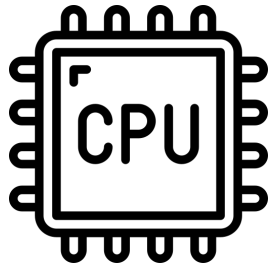


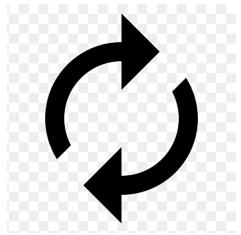
Challenge 2: Resource constrained Data-preprocessing



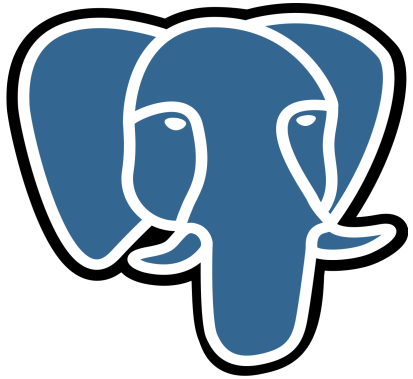


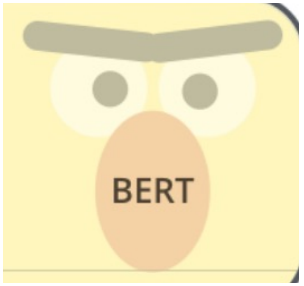


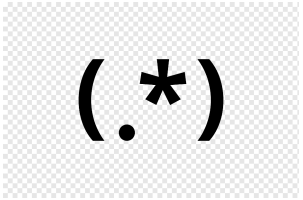


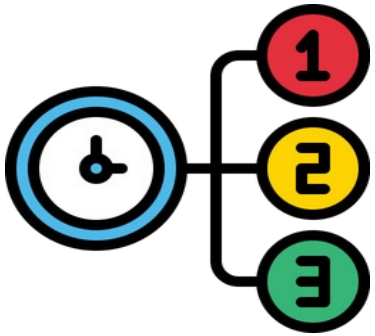


Resource Constrained Feature Extraction









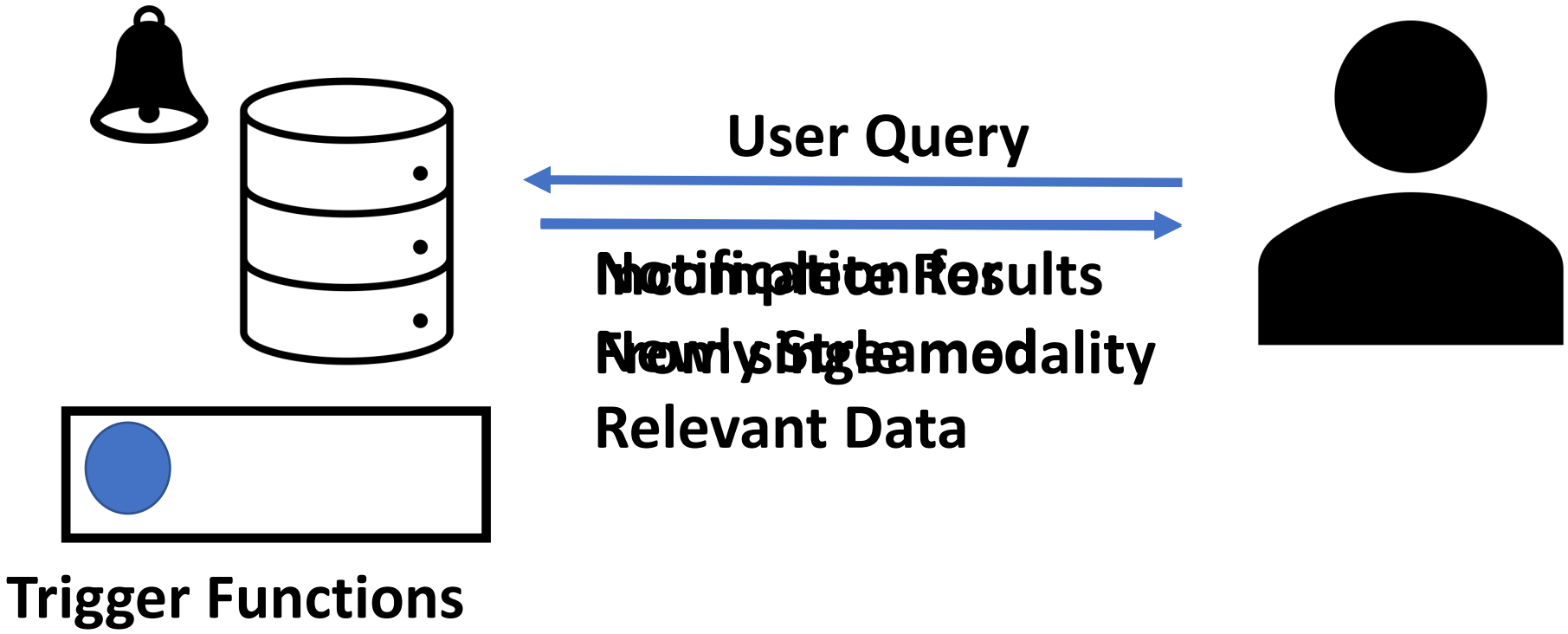
Video and Image Feature Extraction

- Priority System
- Object Detection

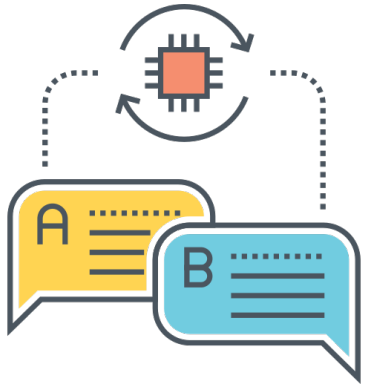
Text Feature Extraction

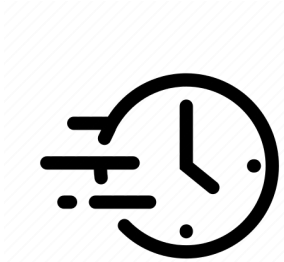
- Regular Expression
- Word Embedding
- Language Models
- PoS-based Classifiers

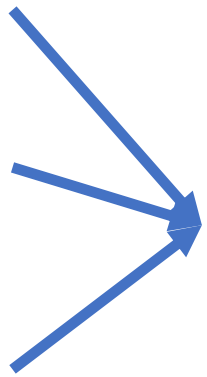
Query results over time

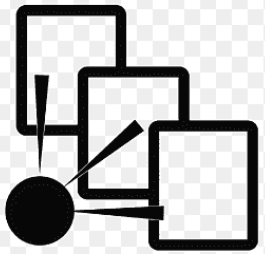


Challenge 3: Label Independent Data Integration

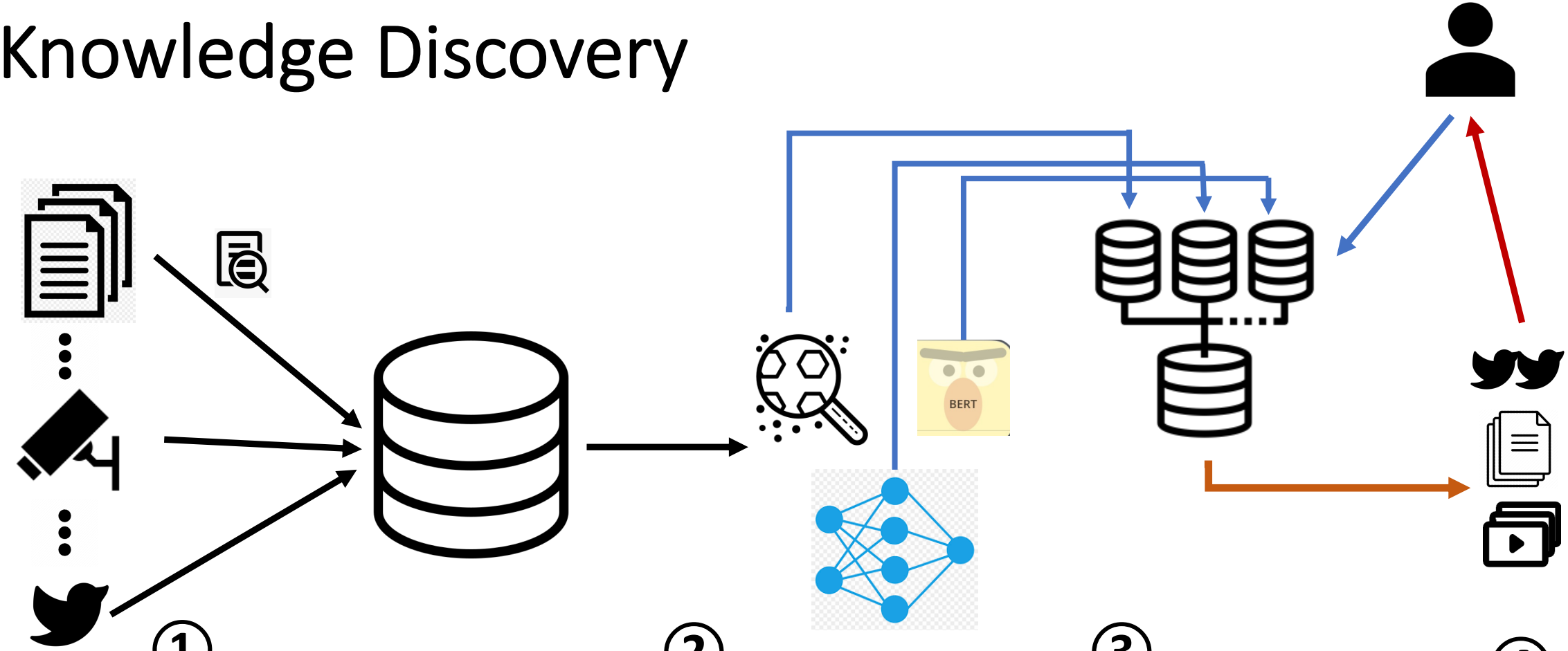








Knowledge Discovery



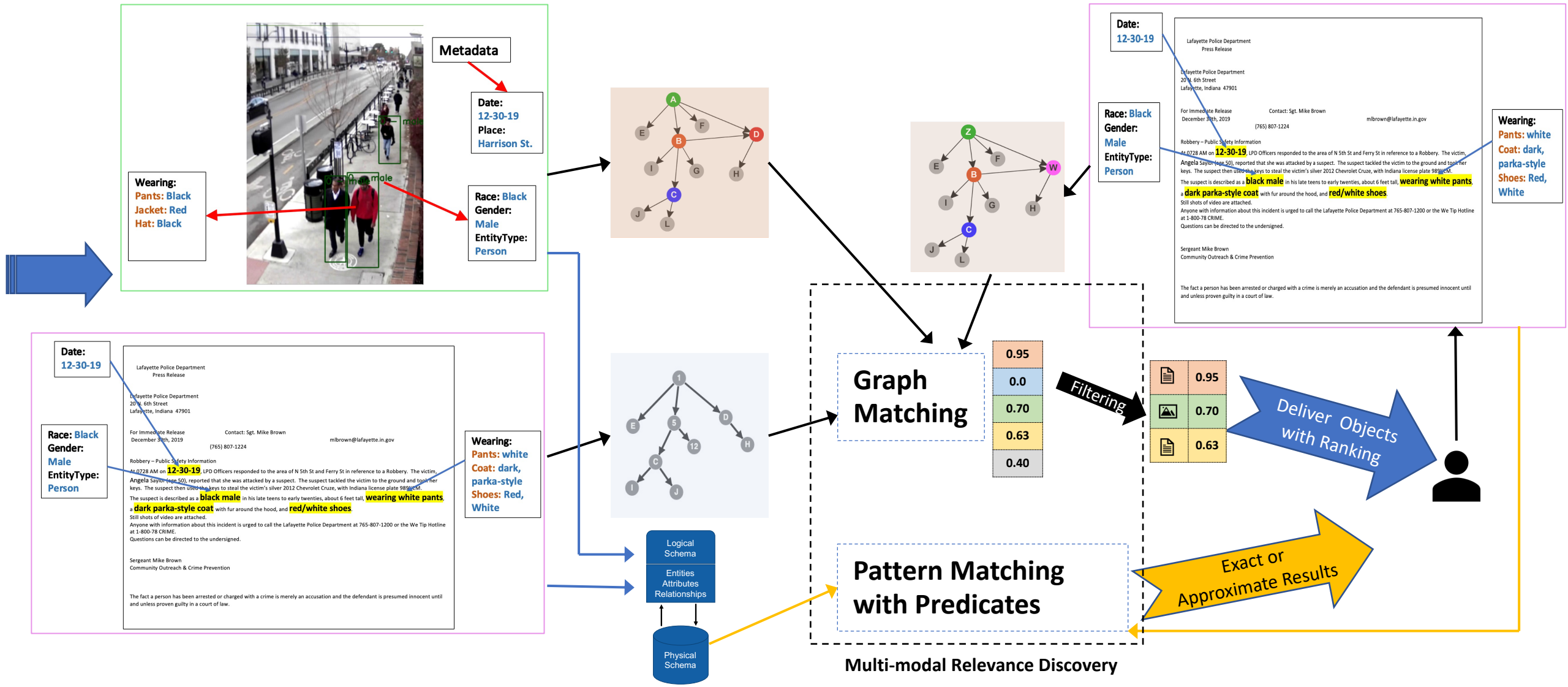
① Capture Streaming Data; Pre-processing; Convert to MP4, txt; Store in Postgres **(Data Capture and Data Storage)**

② Run ML and heuristic models on video, text and tweets **(Feature Identification)**

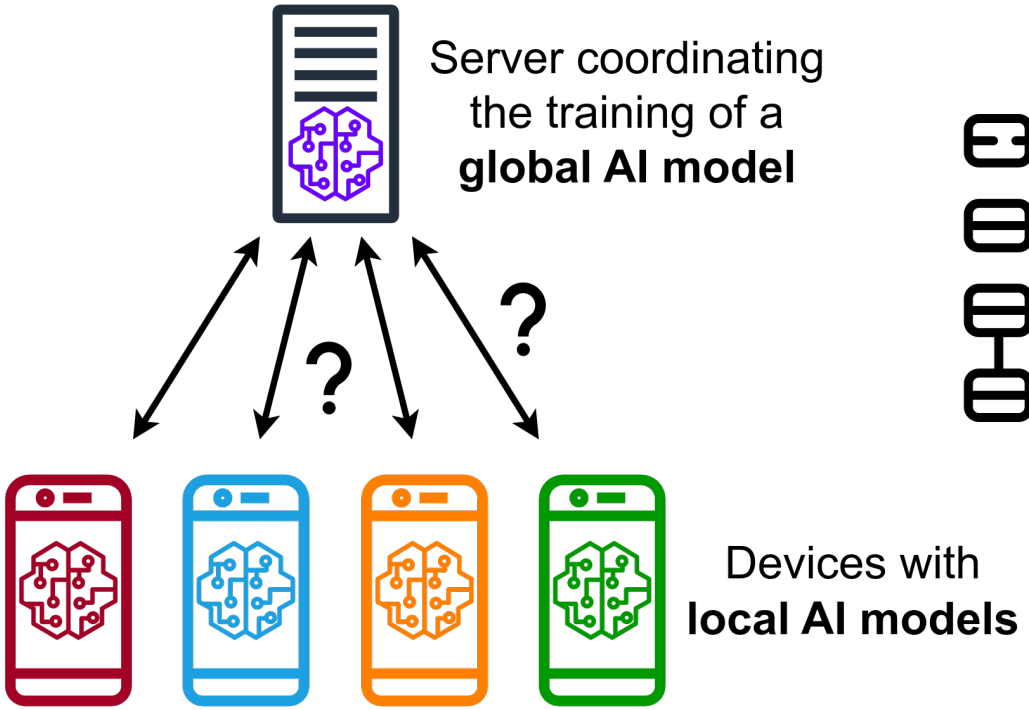
③ Create Query from User Requirement with DDL and return JOIN results **(Data Fusion)**

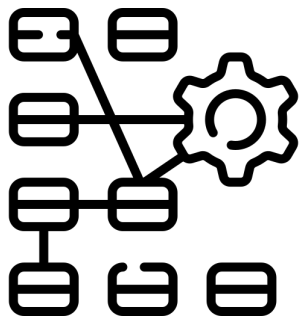
④ Return relevant results to user's examples **(Data Retrieval)**

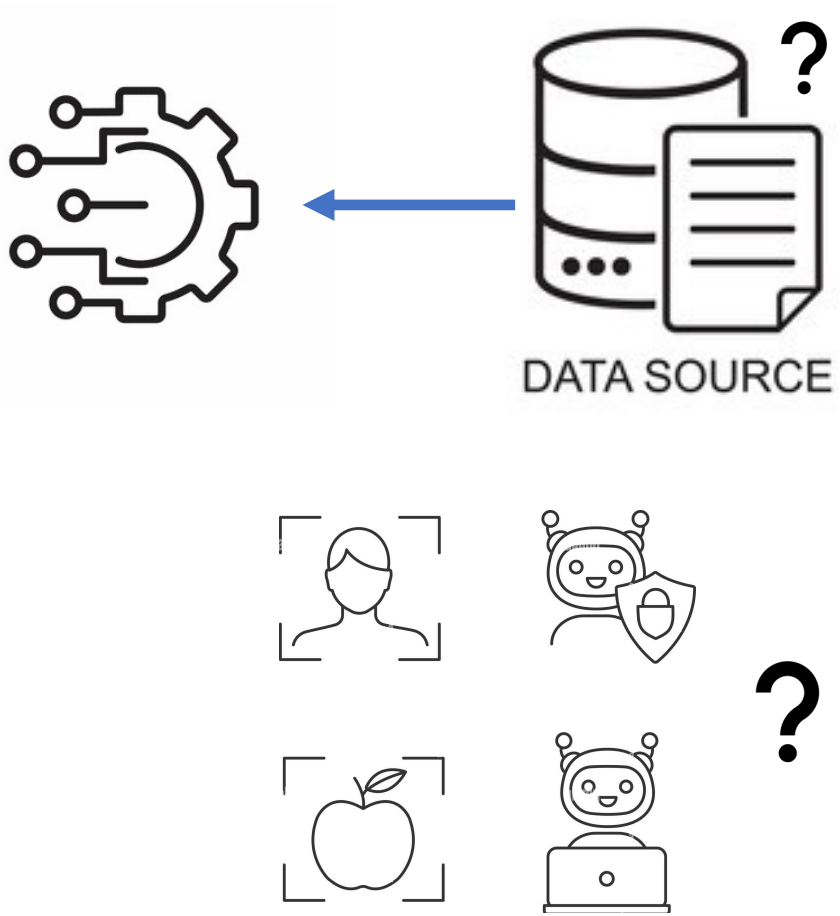
Relevance Modeling and Data Fusion



Challenge 4: Adaption to Open-world Novelties

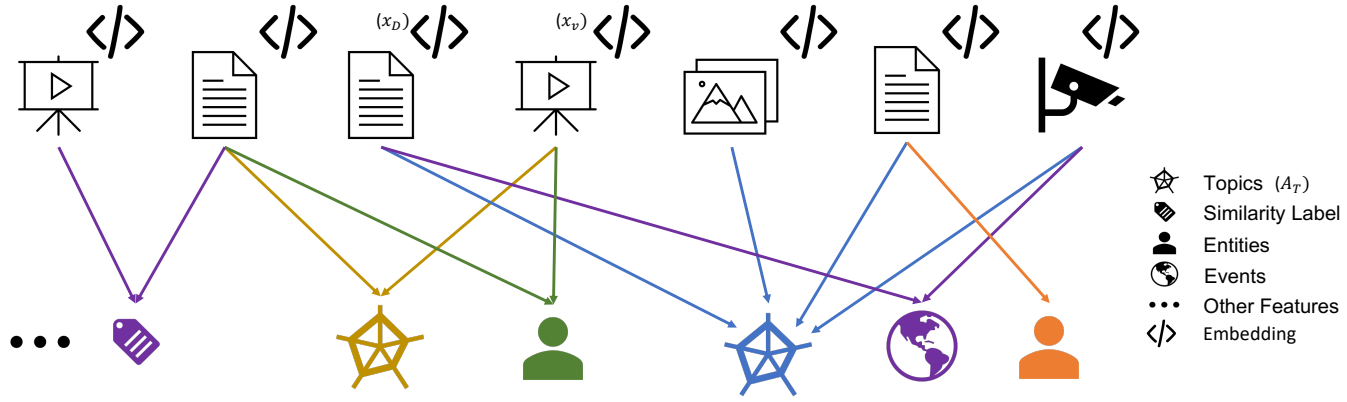






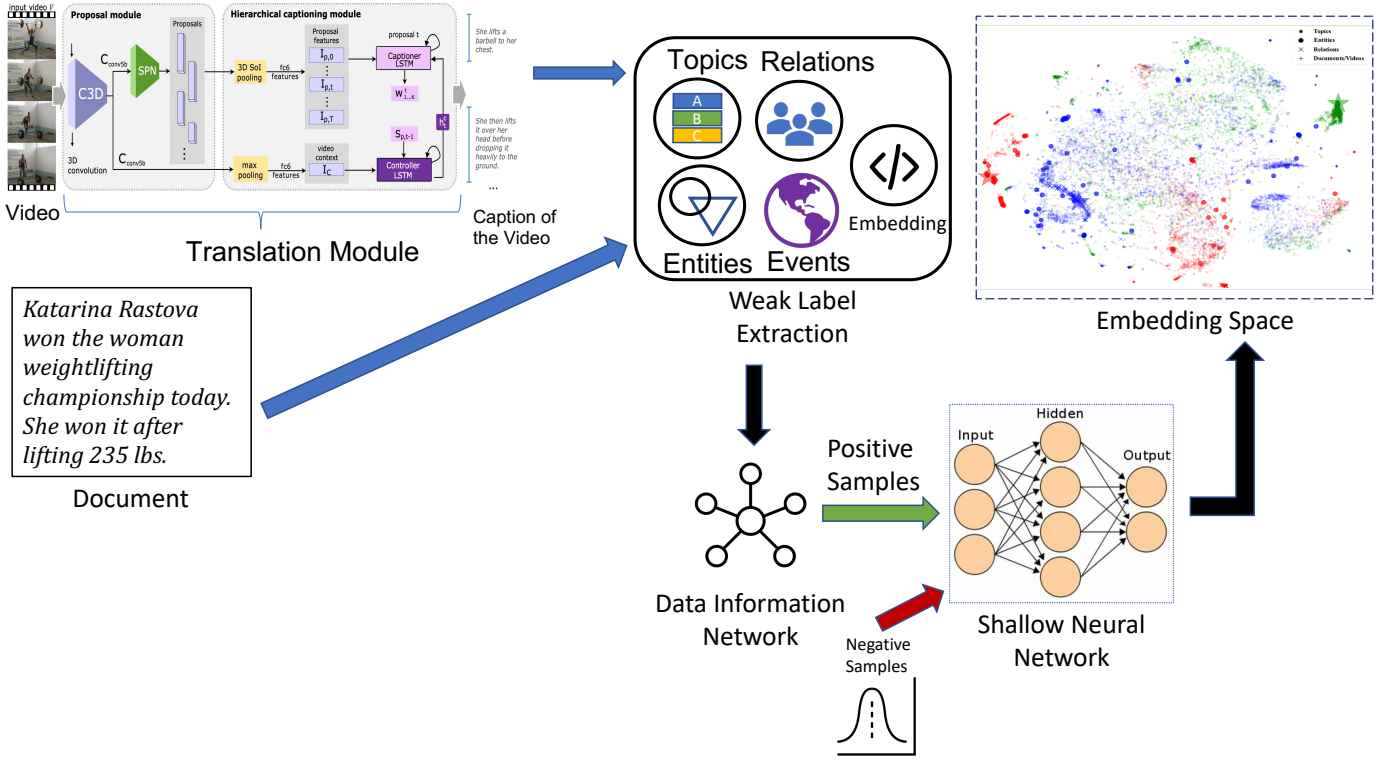
Novelty detection

- ***Data information network*** is used to detect the changes during post-novelty inference.
- **Novel Instance.**
  - A test instance x is novel if $G(V_{P_{tr+x}}, E)$ is different from $G\left(V_{P_{tr}}, E\right)$.
  - Considering a knowledge base for the weak features during training (A_{tr}), if weak features are absent in A_{tr} during testing, the instance is novel.



Weakly Supervised Learning

- translation from different modalities of data to a textual representation
- weak feature labels extraction
- creation of data information network by connecting data samples to their features via their interactions
- construct a structure-infused textual representation, by jointly embedding in a single space
  - the data samples,
  - the features in which these data samples are similar,
  - the similarity labels associated



Similar System in Practice

NORTHROP GRUMMAN

- https://www.fbi.gov/services/cjis/ndex
- Unclassified national information sharing system that enables criminal justice agencies to search, link, analyze, and share local, state, tribal, and federal records.
- Strategic investigative information sharing system that fills informational gaps and provides situational awareness.
- **Analysts: Connecting the Dots**
- **Detectives: Linking Investigations**
- **Patrol Officers: Preparing for Encounters**
- **Regional Dispatchers: Increasing Officer Safety**

- incident, arrest, and booking reports; pretrial investigations; supervised released reports; calls for service; photos; and field contact/identification records.

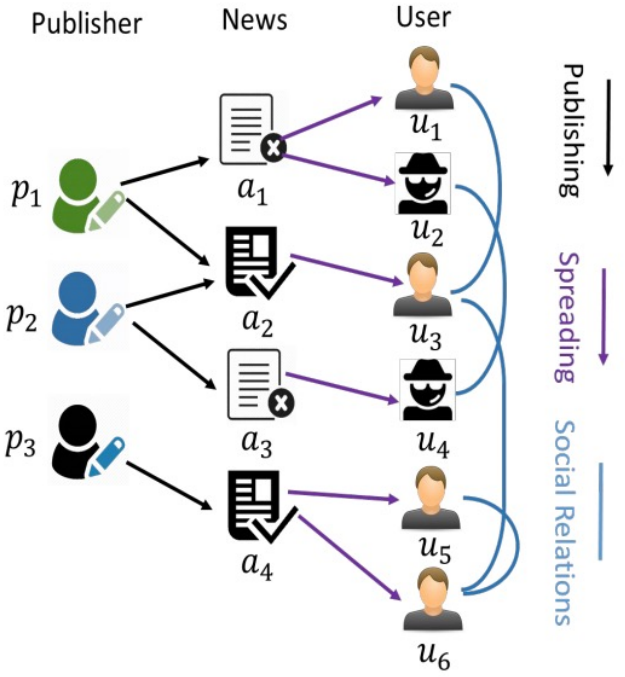
Future Research Directions

Interactive reasoning to cross check integrity and credibility of multimodal data

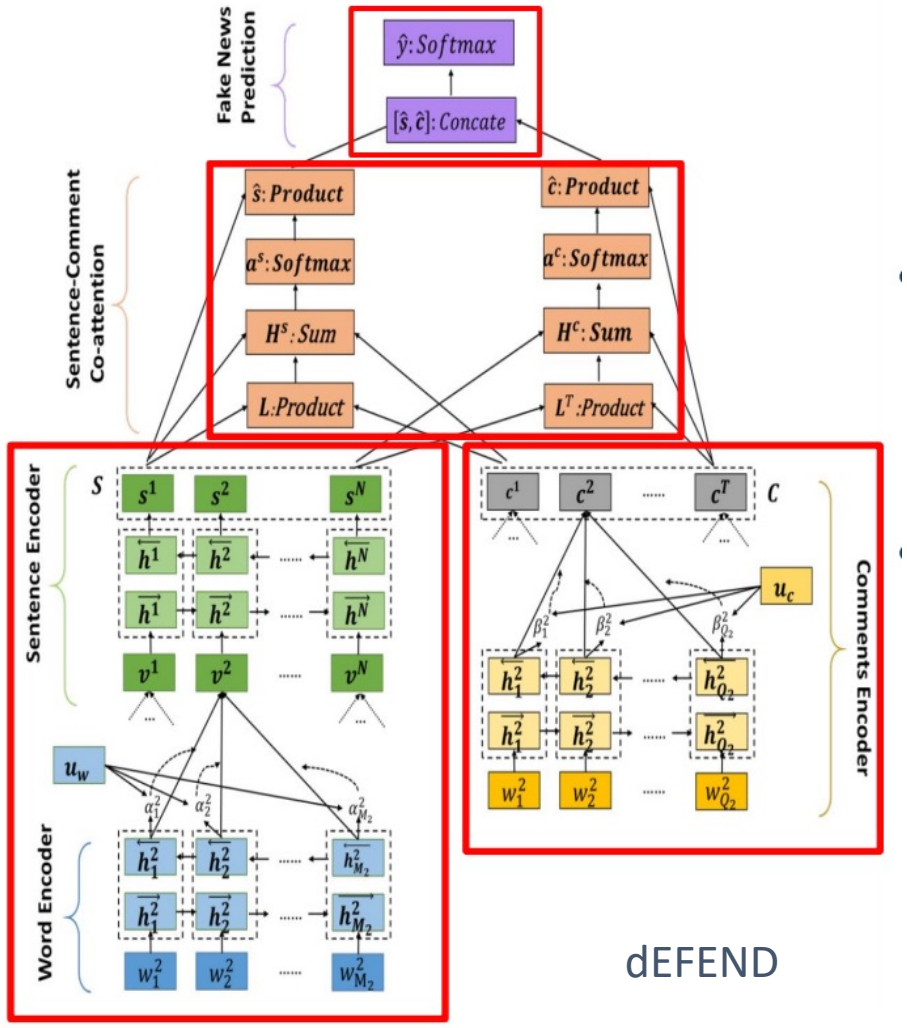
NORTHROP GRUMMAN

- Detecting fake police leads/ tweets/ [report/ tip news articles] and explaining why it is detected as fake
  - Provide insights and knowledge to domain experts
  - Explainable features from noisy auxiliary information can further help detection performance

- Social context provides rich auxiliary information beyond news content [Tweets and Reports]
  - Goal: learn representations from the heterogeneous network
  - Jointly embedding reports/ news articles and social context
- Information from different modality can help to explain and detect authenticity of another [WeTip News and Tweets]
  - How to model content-content relations?
  - How to leverage authentic knowledge base structured information?



Detection of information credibility with explanation



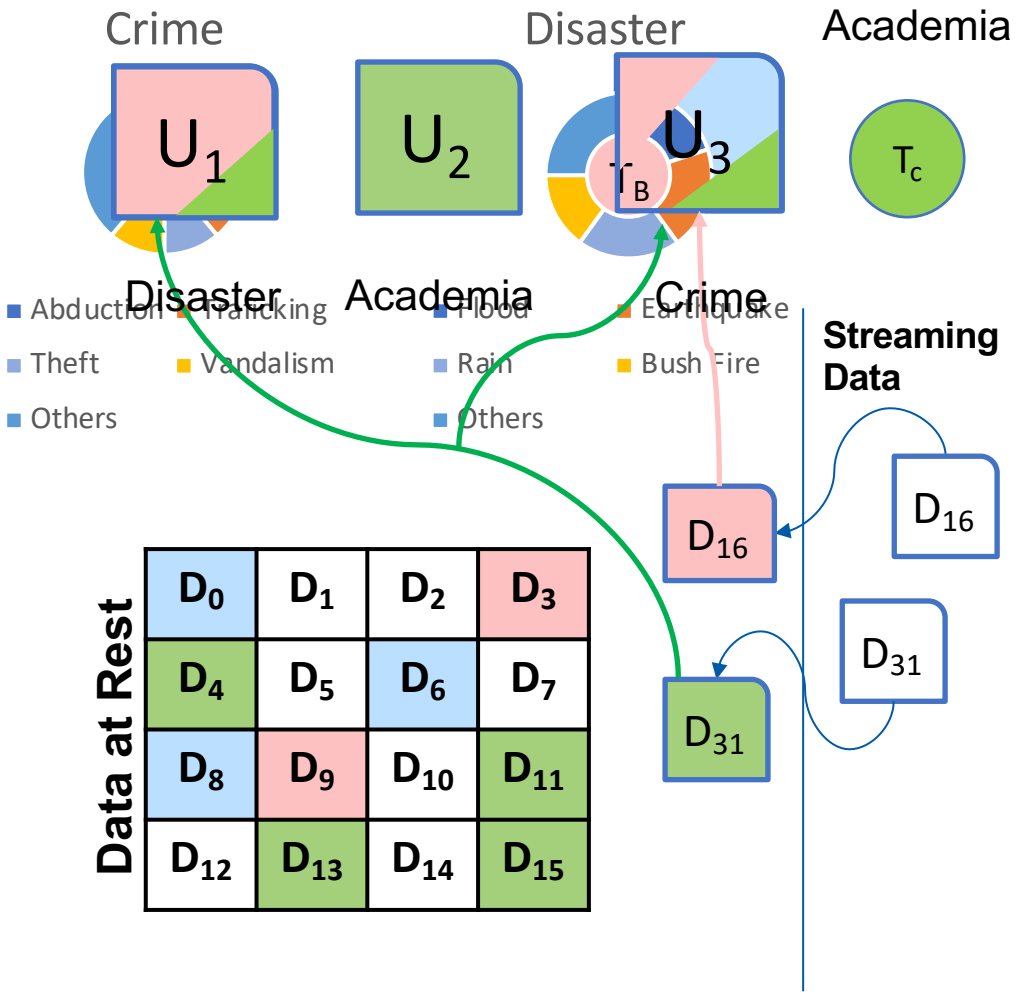
dEFEND

- Learn representations for each modality of data
  - Different Attention Networks depending on the data type
- Select top explainable sentences and tweets through a co-attention network
- Detect fake leads with concatenated sentence and tweet representations as Classification task

Kai Shu, Limeng Cui, Suhang Wang, Dongwon Lee, and Huan Liu. ``dEFEND: Explainable Fake News Detection", KDD 2019, August 4-8, 2019. Anchorage, Alaska.

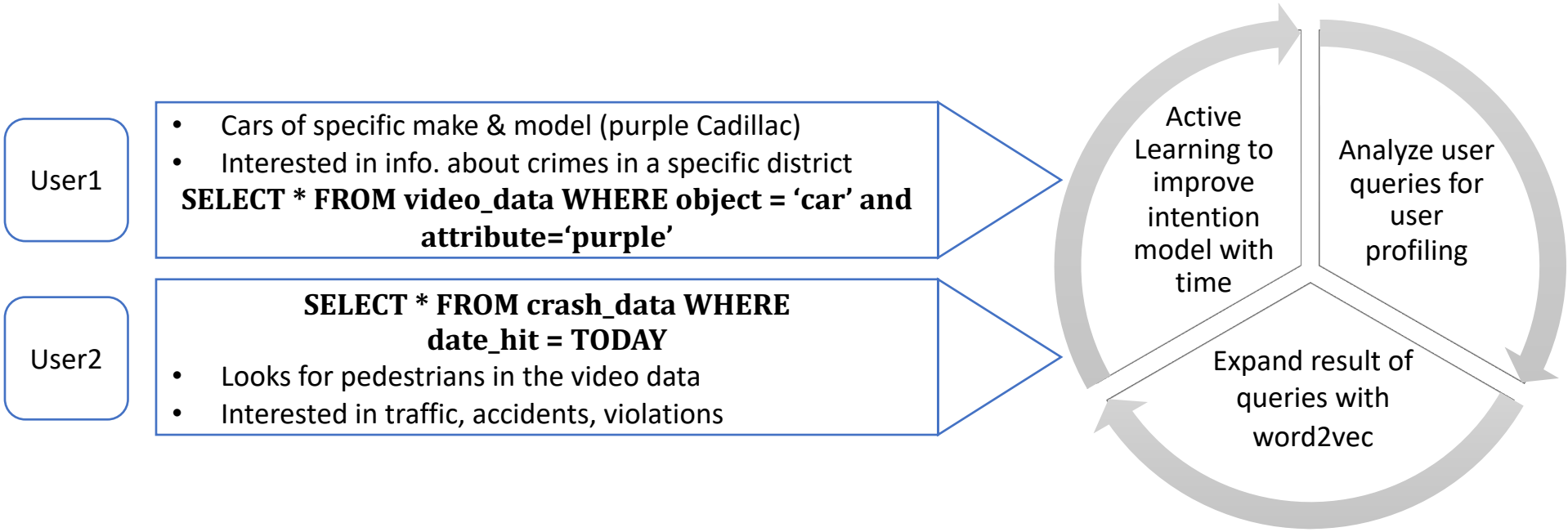
Multiple Data of Interest to Different Users

- Extract **human-interpretable topics** from a document corpus
- Each topic characterized by words most strongly associated with
- Documents as mixtures of topics that spit out words with certain probabilities.
- **No need to re-train**



User Modeling: Intention-aware Recommendation Engine

- Sends users streaming data that corresponds to their interests
- Builds User Profiles using the history of user queries
- Active Learning to narrow/expand intention model with more interaction
- Expands user queries with word embedding models to fetch relevant data from the database



Potential Collaborations

- Existing interdisciplinary research centers and initiatives
  - Institute for Defense Analyses (IDA), Information Sciences Institute (ISI)
    → Novelties in Planning domain
  - MIT (Mike Stonebraker) and University of Michigan (Mike Cafarella)
    → Situational Knowledge on Demand

Collaboration	Area	Collaborators
Explanability and Trust	Multimodal information retrieval	Prof. Yongjian Fu
Resource Management, Information Completeness	Disaster Resilience	Prof. Satish Kumar
Weak supervision, Credibility, User Modeling	Social media analysis and Big Data	Prof. Sun Chung
Scalability and Unsupervised, **information credibility with explanation**	Situational Awareness	Prof. A Essa

THANK YOU

QUESTIONS?

Weakly Supervised labels

- Representing data in terms of different structural features through which different modalities of data can be similar
- Structural representation of raw unstructured texts (with topics, entities, events, and relationships) allows readers to infer better knowledge
- Feature labels are generated automatically in two steps –
  - a textual description of each data sample is generated from any modality;
  - topics, entities, and events are extracted from the textual descriptions and are considered as weak labels for two reasons.
    - quality of the extracted structural units rely on the choice of the extraction models and can be noisy.
    - output generated from the modality specific textual descriptors can be ambiguous and noisy.

Multi-task learning

- For each object, o_i in the graph participating in relation R, s_i^p and s_i^n refers to positive and negative examples. e_{o_i} refers to the vector embedding of the graph object o_i , and y is the label.
- $y = 1$ for (o_i , s_i^p) pairs and $y = 0$ for (o_i , s_i^n)

For each individual graph relation, R, we can define the learning objective as follows:

$$L_R = \sum_i L(o_i, s_i^p, s_i^n) \tag{1}$$

$$L(o_i, s_i^p, s_i^n) = y \log sim(o_i, s_i^p)) + (1 - y) \log(1 - sim(o_i, s_i^n))) \tag{2}$$

where $sim(o_i, s_i^p) = \sigma(e_{o_i} \cdot e_{s_i^p})$;
$sim(o_i, s_i^n) = \sigma(e_{o_i} \cdot e_{s_i^n})$

Learning objectives

- Features to Features ($A_T A_T$ /$A_n A_n$/$A_{event} A_{event}$)
  - Similar topics, named entities, or events with embedding value within a certain threshold, are placed together
- Data Sample to Data Sample ($x_D x_V$ /$x_D x_D$/$x_V x_V$)
  - Positive pairs are selected by
    - Topics, Events and Entities, User Provided similarity labels, and Embedding
- Data Samples to Features ($x A_T$ /$x A_{event}$/$x A_n$)
- Joint Object Function, $L_{total} = \sum_{i \in O_s, O_s \subset O} \lambda_i L_i$

  where O is defined over all the objectives, weight λ_i is set to 1.

$$Rel(a,b) = \frac{\sum_{i \in P(a,b)} w_i}{\sum_{b \in B} \sum_{i \in P(a,b)} w_i}$$

$$Rel(a,b) = sim(e_a, e_b)$$

$$Rel(f,b) = I * N_p(f,b)$$

Reasoning Over the Data Information Network

- Weak Supervised Baseline
  - With the data information graph
  - #paths from one data sample to a given data sample or a given feature
  - counting the paths from one data sample to a given data sample or a given feature.

- *Similarity Based Score.*
  - Given a data sample, or a feature a and their embedding e_a the relevance score with other data sample b with embedding e_b is:

Novelty Characterization

- Covariate shift with change in application domain with the modalities for which translation module is available (covar-1).
- Prior probability shift with novel weak features (prior-1).
- Prior probability shift with no weak features (prior-2).
- Prior probability shift with novel relevance label (prior-3).
- Temporal concept drift with previously relevant data being non-relevant (concept-1).
- Covariate shift with new modality introduction (covar-2).

Novelty response

- pre-trained retrieval model from WeS-Jem
- three level training strategy
- With new modality introduction novelty, both image and LIDAR modality can be handled with the video translation module. Initial text embedding approaches can generate text embedding for any textual input for prior data shift.
- Linear embedding layers in WeS-JEm maps the OOD inputs into the pre-trained joint embedding space
- For (prior-2) novelty, when system relearns, only the (xx-embedding) objective functions remains active
- For novel modality introduction, a new translation method can be learned.
- User similarity labels provided by Relevance Feedback module have greater weights than old ones