# HARD-SOFT DATA FUSION WITH CHATGPT: TOWARD STRUCTURED REPRESENTATIONS AND AUTOMATED REASONING

Nicholas Gahman  and Vinayak Elangovan

Computer Science program, Penn State University Abington, PA, USA

## ABSTRACT

*Hard and soft data fusion is a foundational concept in data science and information fusion, enabling the integration of quantitative (hard) data with qualitative (soft) information to provide richer, more actionable insights. This work investigates multiple strategies for fusing hard and soft data, analyzing their respective strengths and limitations. As a step toward systematic comparison, one such approach is implemented and evaluated. Building on the ChatIE framework by X. Wei et al., this paper introduces a ChatGPT-based extension capable of transforming unstructured natural language into structured data representations. Additionally, it presents an initial prototype of an automatic inference system designed to interpret and act upon the outputs of data fusion processes, laying the groundwork for more advanced decision-support tools.*

## KEYWORDS

*Data Fusion, Graph-based Fusion, Text Transformation, Hard Data, Soft Data*

## 1. INTRODUCTION

In today's data-driven world, effective decision-making across sectors increasingly depends on the ability to interpret and integrate large volumes of diverse data. This challenge is particularly acute in the domain of human activity recognition, where applications such as intelligence analysis and surveillance demand the rapid synthesis of information from heterogeneous sources to inform critical judgments. Central to this challenge is the fusion of two broad categories of data: *hard* data and *soft* data. Hard data is typically quantitative, structured, and derived from physical sensors and measurement devices. It is often reliable and precise, comprising numerical values that represent measurable phenomena. In contrast, soft data is qualitative, unstructured, and subjective—frequently originating from human sources, such as narrative reports, eyewitness accounts, or expert assessments. Despite its inherent ambiguity, soft data carries rich contextual and semantic information that is essential for interpreting complex situations. The integration of hard and soft data—commonly referred to as hard-soft data fusion—has demonstrated value across a wide range of domains, including healthcare, finance, environmental monitoring, and military operations. By combining the objectivity of hard data with the interpretive depth of soft data, fusion techniques can support more comprehensive analysis and enable more informed, context-sensitive decision-making.

A growing number of algorithms and frameworks have been proposed for hard-soft data fusion, each with distinct advantages and trade-offs. However, identifying the most effective approach for a given application remains an open question. To contribute to this area of research, the present work explores several common fusion methodologies, implements one selected approach,

and evaluates its performance using a unified dataset. In doing so, it aims to advance practical understanding of how these fusion strategies operate and how they can be further optimized.

This study also extends the ChatIE framework proposed by X. Wei et al. by developing a ChatGPT-based system for converting unstructured textual input into structured representations. Additionally, a preliminary automatic inference system is introduced to support reasoning over the fused data. Together, these components provide an early foundation for an end-to-end system capable of transforming raw multimodal data into actionable insights.

## 2. RELATED WORKS

The integration of hard and soft data has garnered significant attention in recent years, particularly in domains such as intelligence analysis, military operations, and situational awareness. Numerous studies have proposed frameworks and algorithms to address the challenges of fusing heterogeneous data sources for improved decision-making.

Gross et al. [1] presented a framework for hard-soft data fusion in the context of intelligence analysis. Their approach involves processing soft data using natural language processing (NLP) to generate propositional graphs that represent entities, events, and relationships. These graphs are then transformed into attributed graphs that incorporate uncertainty, including observational bias and descriptive variance. Hard data is similarly converted into attributed graphs, and the two datasets are fused by merging identical portions and resolving conflicts based on uncertainty estimates. The fused graph is subsequently analyzed using graph matching techniques to identify situations of interest. The soft data processing is accomplished using the Tractor system, which employs the General Architecture for Text Engineering (GATE) to extract syntactic information, followed by mapping to semantic structures. Notably, the authors prioritize modifying hard data to align with the less structured soft data to avoid introducing artificial precision. Date et al. [2] addressed data association in counterinsurgency operations, with a focus on identifying overlapping information between hard and soft sensor observations. Their methodology converts data into relational attributed graphs and uses feature-based similarity scoring to compare graph components. Nodes and edges with high similarity are clustered and merged to form a unified data graph. The study also introduces an evaluation framework based on precision, recall, F-score, and computational efficiency. While the evaluation primarily targets soft-soft data association, the authors acknowledge the need to extend the framework to support hard-soft and hard-hard data fusion.

Llinas et al. [3] proposed a generalized hard-soft data fusion architecture for intelligence applications, emphasizing modularity and adaptability. Hard data is sourced from sensors such as radio frequency systems and satellite imagery, while soft data is obtained from unstructured operational reports. The authors advocate for separate initial processing of hard and soft data, followed by late-stage fusion. Their framework supports multi-source fusion, including human-generated reports and web-based information. Although the methodology is still under development, the framework integrates semantic search algorithms and temporal graph representations to enhance the analysis of evolving scenarios. Additionally, the study explores the use of conceptual spaces—structured representations of human concepts—as a promising mechanism for improving fusion capabilities, particularly in handling linguistic data. Chapman et al. [4] build upon the conceptual space framework introduced by Llinas et al., applying it to the prediction of kinetic kill space events in satellite operations. Their system models both hard and soft data within ontologies, which provide a standardized and hierarchical structure for representing domain knowledge. Conceptual spaces are defined through quality dimensions, domains, and properties, enabling the representation of abstract entities and relationships. The authors develop a model to estimate the likelihood of hostile interactions between spacecraft by

analyzing four factors: intent, opportunity, capability, and vulnerability. Each factor is quantitatively represented using conceptual space models, which are then integrated to support predictive reasoning.

Ahmed et al. [5] explored hard-soft data fusion through the lens of Bayesian state estimation. Their method is designed to manage uncertainty in soft data, which is expressed in structured linguistic templates. The proposed framework enables cooperative search and identification tasks involving both a human operator and an autonomous robot. The system models soft information as probabilistic constraints and demonstrates improved performance when the robot and human collaborate. The study emphasizes the need for application-dependent interpreters to preprocess soft data, drawing parallels with preprocessing techniques for sensor data such as LiDAR or imaging. Reece et al. [6] introduced a kernel-based approach to data fusion, leveraging feature vectors that encapsulate both hard and soft data. These vectors are assumed to be complete and homogeneous in format, allowing the use of transformation kernels to enable comparison between heterogeneous features. Gaussian Processes (GPs) are used as the core classification mechanism, offering a probabilistic interpretation of data relevance and classification outcomes. The system is applied to a simulated scenario involving improvised explosive device (IED) detection, where it successfully identifies relevant data sources and distinguishes between benign and hostile events. Although the soft data in this study is structured using rank and nominal kernels, the authors note the ongoing challenge of interpreting unstructured soft data for Bayesian fusion models.

K. Vasnier et al [7] proposed Dynamic Bayesian Networks (DBNs) for real-time situational awareness in crises. DBNs model time as discrete slices with variable dependencies captured in a directed graph, enabling hierarchical reasoning from sensors $\rightarrow$ observables $\rightarrow$ inferable $\rightarrow$ hypotheses. The system identifies which variables are critical for reliable hypothesis inference and then selects the most relevant sensors accordingly. This top-down information acquisition allows for fast and efficient decision-making. However, the approach suffers from exponential complexity as sensor action choices increase. Future work involves designing heuristics that manage this complexity and account for sensor action costs (e.g., time, energy). J. R. Chapman et al [8] compared Conceptual Spaces with Gaussian noise to Dempster-Shafer (DS) theory for handling soft data uncertainty. Conceptual spaces need similarity metrics to relate observations to conceptual categories. The previous state-of-the-art method used normalized distance metrics to determine similarity, while DS theory constructs a Frame of Discernment (FoD) and assigns belief and plausibility values to subsets. In experiments on space-domain classification (satellite, aircraft, rocket body), DS theory consistently outperforms the normalized distance metric in similarity scoring and uniquely supports conflict detection between sensor inputs. The study concludes that DS-based similarity metrics yield better accuracy and interpretability for uncertain data fusion in conceptual spaces.

T. L. Wickramarathne et al [9] introduced the DS Conditional Approach to overcome limitations of fusing soft data from differing Frames of Discernment (FoDs). Traditional Evidence Fusion assumes a common FoD and struggles with ambiguous or contradictory data. The authors develop the Conditional Update Equation and Conditional Core Theorem, which enable belief function fusion across differing FoDs without modifying the original frames. The resulting approach is both more accurate and computationally efficient, achieving up to 80% faster performance in practical settings. S. Acharya et al [10] built on DS theory to handle contradictory soft data sources by introducing the Consensus Operator, which adds a fourth element—relative atomicity—to the standard belief, plausibility, and uncertainty framework. Relative atomicity captures how granularly aligned an observation is with other evidence. This enables computation of a collective probability expectation across all soft sources, which can then be fused with hard sensor data using traditional probabilistic methods. This method improves robustness in the

presence of conflict without requiring redefinition of FoDs. X. Wei et al [11] proposed ChatIE, a two-stage zero-shot information extraction system utilizing ChatGPT's strength in question answering. Instead of extracting information in one step, ChatIE first asks what types of relations, events, or entities are present (from a predefined list), then performs targeted extraction for each type. This decomposition significantly improves performance. ChatIE outperforms existing zero-shot IE methods and even surpasses several fully supervised models, demonstrating that large language models can generalize to structured extraction tasks when properly prompted.

## 3. METHODOLOGY

Building on the reviewed literature, we identify the most effective data fusion techniques, which can be broadly classified by their focus: event-based or entity-based. Since soft data often lacks reliable spatial and temporal information, we prioritize event and entity-focused methods. The three most promising techniques were chosen due to their consistent effectiveness, versatility and independence from external formats such as ontologies:

- Bayesian Fusion (event-based)
- Dempster-Shafer Evidence Fusion (event-based)
- Graph-based Fusion (entity-based)

**Bayesian Fusion:** Bayesian Fusion begins by defining a set of hypotheses—possible events to be inferred. A probabilistic model is constructed that includes prior probabilities for these hypotheses and the statistical relationships between input attributes. As new data sources are introduced, Bayes' Theorem is applied to update the probability of each hypothesis, allowing for selection of the most probable outcome. Figure 1 shows an overview of Bayesian Fusion's process.
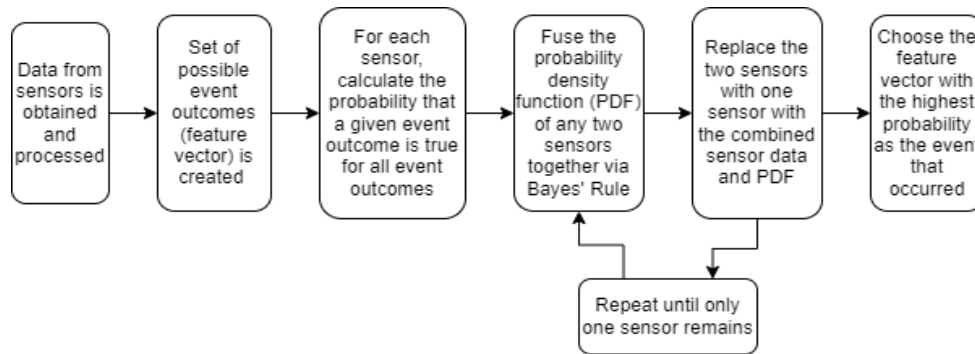


Figure 1: High-level Diagram of Bayesian Fusion

However, Bayesian Fusion has notable limitations:

1. Assumption of Data Independence: It assumes all data sources are independent. In soft data scenarios, this is often invalid due to hidden dependencies (e.g., coordinated sources), leading to skewed results.

2. Scalability Issues: As the number of parameters grows, so does the computational burden of updating the model with new data, making it less practical for large-scale applications.

3. Sensitivity to Priors: The model's output is heavily influenced by its initial prior probabilities, which are often difficult to estimate accurately. Poor priors can degrade performance and require careful tuning for each scenario.

**Evidence Fusion (Dempster-Shafer Theory):** Evidence Fusion begins by defining a Frame of Discernment (FoD)—a set of all possible hypotheses, where multiple hypotheses can be true simultaneously, unlike in Bayesian Fusion. For each data source, a Basic Probability Assignment (BPA) is calculated, representing the probability that a given subset of hypotheses is true.
From the BPAs, two functions are derived:

- Belief: the lower bound of the probability, summing BPAs of all subsets within a given set.
- Plausibility: the upper bound, summing BPAs of all supersets of the set.

The gap between belief and plausibility represents uncertainty. These functions are then combined using Dempster's Rule of Combination, typically by fusing two or three sources at a time. When more than three sources are present, the fusion is performed iteratively, treating fused outputs as new pseudo-sources. The subset with the highest belief is taken as the most probable event. Figure 2 shows an overview of Evidence Fusion's process.
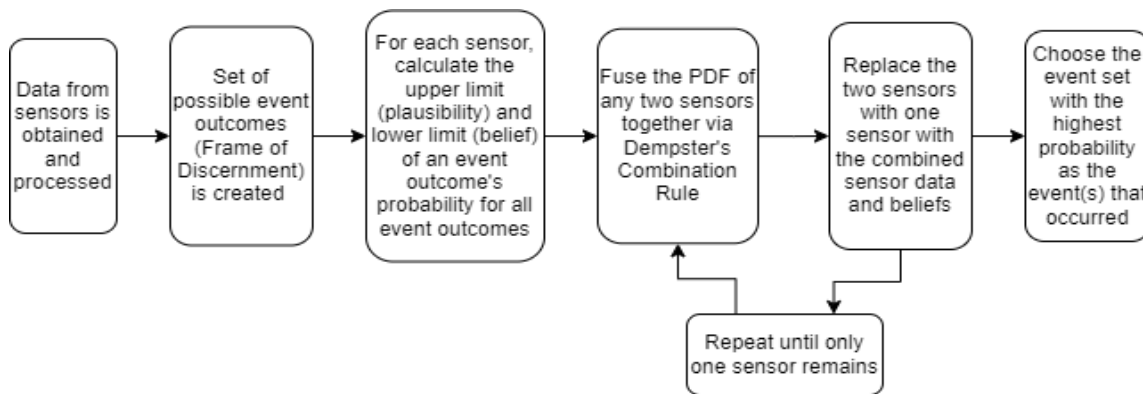


Figure 2: High-level Diagram of Evidence Fusion

Limitations of Evidence Fusion:

1. Scalability with Hypotheses: The number of possible subsets grows exponentially with the number of hypotheses, as many are not mutually exclusive. Each source must assign probabilities to every subset, making the process computationally intensive in complex scenarios.

2. Fusion Complexity with Many Sources: While two-source fusion is manageable, fusing three or more sources becomes increasingly complex. Iterative fusion using pseudo-sources may obscure the conflict between individual sensors, reducing transparency and potentially masking uncertainty.

3. Hypothesis Generation Challenge: Unlike Bayesian Fusion, the challenge lies in constructing an appropriate set of hypotheses. Too few hypotheses limit event granularity; too many inflate computational demands. Hypothesis sets must be carefully tailored to each application.

**Graph-based Fusion:** Graph-based Fusion begins by constructing a graph for each data source, where entities are represented as nodes and their relationships as edges. The fusion process then matches and merges similar nodes across graphs, consolidating data from different sources. Once merging is complete, the fused graph is passed to a neural network or other predictive model to infer the most likely hypothesis or outcome. Figure 3 shows an overview of Graph-based Fusion's process.
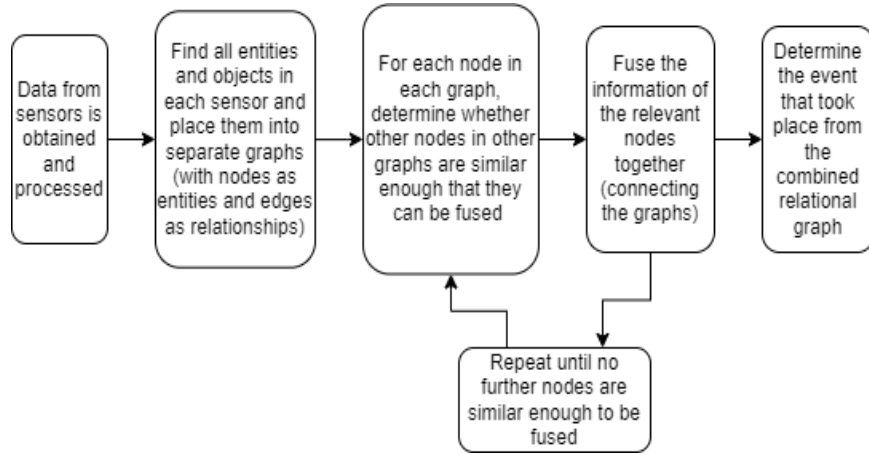


Figure 3: High-level Diagram of Graph-based Fusion

Limitations of Graph-based Fusion:

1. Graph Size and Scalability: The main scalability issue lies in the size of the resulting graph, not in the number of hypotheses or data sources. As more entities are added, matching new nodes to existing ones becomes computationally expensive, especially in real-time settings. Although duplicate entities are merged, the graph still grows if new entities are introduced.

2. Indirect Event Prediction: Unlike Bayesian or Evidence Fusion, Graph-based Fusion does not directly infer events. Instead, it models relationships and requires an external predictive algorithm (e.g., a neural network) to analyze the fused graph and predict events, introducing an extra processing step.

3. Complex Situations: For highly dynamic or dense interaction scenarios, the graph may become too complex to update efficiently, limiting the approach's effectiveness in real-time or large-scale deployments.

To implement and evaluate any of the proposed data fusion techniques, a suitable dataset is essential. However, no publicly available datasets specifically designed for data fusion currently exist. As a result, a custom dataset was created for this project. This was accomplished by analyzing four publicly available videos on YouTube, from which both hard and soft data were extracted. The hard data was formatted using the Transducer Markup Language (TML) standard [12]. A sample structure of this format is shown below:

*<data_ref="source">*
  *month/day/year,   hh:min:sec,   message-id,   Global_Space,   Source_attributes,*
*Local_Space,  Reference_Space,  Detection_Focus,  ID,  Detection_Focus_Attributes,  Group-ID,*
*Detection_Confidence*
  *</data_ref="source">*

Graph-based Fusion was chosen as the ideal technique to implement due to the nature of the custom dataset. Bayesian Fusion and Evidence Fusion both become increasingly difficult to implement as the technique needs to choose between a greater series of outcomes. The custom dataset's videos each show different situations and have different outcomes, so Bayesian and Evidence Fusion's implementation would be very complex. Graph-based Fusion only becomes more complex as the situation itself does, and the custom dataset features simple cases with a small number of entities and events, which makes Graph-based Fusion perfect for the custom dataset. Furthermore, Graph-based Fusion is intuitive, so humans can easily understand the results of the fusion. These reasons make Graph-based Fusion the best technique for implementation.

## 4. IMPLEMENTATION AND DATA GENERATION

With the overall comparison framework established, we now detail the implementation process and dataset creation. A critical aspect of this project involves determining how hard and soft data are generated, as data quality directly impacts the success of fusion.

### 4.1. Hard Data Generation

To generate hard data from a given video, we decompose it into a sequence of image frames. Each image is processed by algorithms designed to extract events and entities, classifying them according to the TML (Transducer Markup Language) format. For this study, we assume ideal algorithm performance—i.e., no misclassifications—since the actual development of classification models lies outside the project's scope. As a result, hard data for four videos was manually curated to ensure consistency. Additionally, only visual data from the videos was used for curation.

### 4.2. Soft Data Generation

Soft data was generated by manually composing textual descriptions of the videos as the authors viewed them. These narrative summaries aimed to capture all relevant events and entities described or implied in the visual content.

#### 4.2.1. Data Integration Strategy

Because hard and soft data inherently use different formats—structured versus unstructured—fusion requires converting both into a common representation. This can be done by either:

1. Structuring the soft data, or
2. De-structuring the hard data (i.e., converting it to text).

Each approach has trade-offs. Structuring soft data enables efficient storage and retrieval of entities, events, and spatiotemporal relationships. However, it risks information loss if certain elements of the narrative don't map cleanly to the target schema. Moreover, structured formats may vary across hard data sources, requiring harmonization prior to fusion. By contrast,

converting hard data into unstructured text allows the use of NLP tools such as text summarization. However, this direction is less explored: most work in this area focuses on extracting structured information from text, not the reverse. Structured data remains preferred due to its precision and comparability. Given the requirements of Bayesian Fusion, Evidence Fusion, and Graph-based Fusion—which all operate on structured inputs—this project adopts the first strategy: converting soft data into a structured format.

### 4.2.2. Extracting Structure from Soft Data

Extracting structured representations (entities, events, and relations) from soft data poses several challenges. The text may vary widely in style and content, and conventional Named Entity Recognition (NER) techniques often fall short, especially when dealing with unnamed or novel entities.

Initial exploration includes:

- Pre-trained deep learning models for NER, which proved ineffective at identifying unnamed entities.
- Rule-based systems, which lack flexibility and showed poor generalization across different texts.

These limitations highlighted the need for a more adaptable solution. Training a model specifically for the dataset was unfeasible, as there was too little data to work with. Eventually, ChatGPT and other Large Language Models (LLMs) were identified as a promising alternative due to their cross-domain capabilities and prior success in information extraction, as demonstrated by the ChatIE framework [11].

### 4.2.3. ChatIE and Custom Extensions

ChatGPT, a large transformer-based model, leverages tokenization, multi-head attention, and contextual embeddings to predict the next output token. Its large-scale training enables it to capture nuanced relationships across diverse domains, making it well-suited for our task. Although ChatGPT-based solutions are well-known in the literature, they were initially overlooked due to significantly underperforming compared to state-of-the-art (SoTA) zero-shot information extraction (IE) methods [13]. However, ChatIE stands out among ChatGPT-based IE systems by decomposing the complex IE task into a multi-stage question-answering (QA) process—an approach that aligns well with ChatGPT's strengths and leads to improved performance [11].

The authors of ChatIE provided a GitHub repository implementing their two-stage approach:

1. Stage 1 - Entity Type Identification: ChatGPT identifies entity types present in the text (e.g., "Person", "Location", "Organization", "Miscellaneous").
2. Stage 2 - Entity Extraction: For each identified type, ChatGPT extracts relevant entities.

We extend ChatIE with two modifications:

- Entity Types: Added categories—"Object", "Vehicle", "Date", and "Time"—to align with TML requirements.
- Two Additional Stages:

- Stage 3 – Entity-Actions Relationship: For each entity identified, ChatGPT is prompted to explain the actions the entity performs, formatted in a simplified TML structure.
- Stage 4 - : This output is further refined into complete TML messages suitable for fusion.

These extensions ensure that both soft and hard data conform to the same structural schema, facilitating direct comparison and fusion across methods.

The specific TML format ChatGPT uses is as follows:
["Date/Time Period", "Time", "Location", "Entity Type", "Entity_Name",
"Entity_Type_Specific_Attributes", "Group_ID", "Confidence"]

"Entity_Type_Specific_Attributes" in this format is a list of attributes specific to the entity type. Each of these attributes will be examined in detail during the explanation of how ChatGPT handles the four processing stages. This uniform format significantly simplifies data fusion by ensuring both hard and soft data share a common structure, requiring only minimal preprocessing. The fourth stage builds on the third by addressing differences between soft and hard TML data. Specifically, hard data uses discrete values for certain attributes, whereas soft data may present more variability. To reconcile this, ChatGPT is asked to classify its output according to the hard data's predefined attribute values, which are then substituted accordingly.

Due to the potential unreliability of ChatGPT's responses, it is necessary to detail how it processes each of the four stages. Since OpenAI provides only general descriptions of ChatGPT's functioning, insights were gathered by prompting ChatGPT directly during each stage.

In Stage One, ChatGPT identifies entity types using a combination of keyword matching (e.g., "man" for Person, "pizza" for Object) and contextual reasoning (e.g., inferring "house" as a Location without a specific address). In Stage Two, it detects entities of the identified types by finding actors involved in actions and classifying them through lexical similarity (e.g., "man" implies a male Person) and contextual cues (e.g., performing actions like walking or speaking suggests a human agent).

Stage Three is more complex, as each entity type has a different set of attributes. However, common attributes across all entities include: "Date/Time Period", "Time", "Location", "Entity Type", "Entity_Name", "Group_ID", and "Confidence". ChatGPT analyzes the sentence in which an action occurs to generate a TML message. If an attribute is not present, it is marked as "None".

- "Entity Type" and "Entity_Name" are provided in the prompt.
- "Date/Time Period", "Time", and "Location" are extracted from temporal and spatial references, with general times like "winter" categorized under "Date/Time Period" and more specific ones like "afternoon" under "Time".
- For "Group_ID", ChatGPT checks for mentions of group membership and assigns an ID if a group exists, otherwise defaults to 0.
- "Confidence" is determined based on sentence ambiguity, ranging from 100 (clear and unambiguous) to 0 (highly ambiguous).

Additional attributes vary by entity type:

Person-specific attributes:

- "Human Clothing" and "Social Role" are extracted from descriptive elements (e.g., "red shirt" → Clothing; "drives the car" → Driver → Social Role).
- "Human Posture" is inferred from implied physical stance based on actions (e.g., walking implies standing).
- "Human Motion" captures actions not involving interaction with others (e.g., walking, opening a trunk).
- "Human_Human Interaction" includes actions involving other people (e.g., talking).
- "Vehicle_ID" identifies the vehicle most interacted with in the sentence.

Object-specific attributes are ["Object Color", "Object Type", "Object Size", "Object Shape", "Human-Object Interaction", "Human_Name"]

- "Object Color" is derived directly from the sentence.
- "Object Type", "Size", and "Shape" are inferred using ChatGPT's trained world knowledge.
- "Human-Object Interaction" and "Human_Name" are extracted from interactions described in the text (e.g., "delivered by Bob").

Vehicle-specific attributes are ["Vehicle Color", "Vehicle Type", "Vehicle Speed", "Motion Direction", "Vehicle State", "Human-Vehicle Interaction"]

- Most are directly inferred from the sentence.
- "Vehicle Type" is matched against known types (e.g., "car", "bus").
- "Vehicle State" describes whether the vehicle is "parked", "moving", etc.
- "Human-Vehicle Interaction" captures actions involving both entities (e.g., "opening trunk").

TML defines certain interaction attributes—such as "Human Posture", "Human Motion", "Human_Human Interaction", "Human_Object Interaction", and "Human_Vehicle Interaction"—more narrowly than ChatGPT's interpretations. For instance, "Human Motion" in TML refers to general movement types like "Walking", "Pushing", or "Stationary", not specific interactions like "throwing pizza". Additionally, ChatGPT may assign multiple actions to a single attribute, complicating comparisons. Thus, a fourth stage is needed.

In Stage Four, ChatGPT's outputs are post-processed by reclassifying these attributes into standardized TML interaction attributes as specified earlier, to enable consistent comparison. For interaction-based attributes, TML's predefined categories are supplemented by a set from "Group_Activity Type" (e.g., "Object_Delivery", "Vehicle_Changing") to increase semantic richness. While "Human Posture" generally aligns well, other attributes require deeper classification:

- For "Human Motion", ChatGPT maps actions to broader movement categories by analyzing their physical properties (e.g., "opens trunk" → Pushing).
- Similar reasoning applies to the three types of human interactions, with actions conceptually matched to the closest TML category.
- In cases with multiple actions, ChatGPT assesses their aggregate similarity to the available categories and selects the most representative one.

Graph-based Fusion was selected for implementation due to the reasons mentioned in Section 3. Each TML message is distilled into a graph comprising nodes representing the entity, its location, group, and action at a specific time. Entities are prefixed to denote type: "H" (Person), "O" (Object), and "V" (Vehicle). Each message contributes edges connecting the entity to its associated nodes, creating a fully connected representation. The current system builds two separate graphs—one for hard data and one for soft data. Though the next step would be to merge them by matching node attributes, this was not completed due to time constraints. A basic automatic query answering system was also developed. It uses Breadth-First Search (BFS) to identify entities related to a given node.

- A direct relation exists if there is a path from the node to an entity that doesn't pass through another entity node or any activity-activity relationship.
- An indirect relation involves a path that passes through exactly one other entity node or through at least one activity-activity relationship.

This provides a foundational step toward a more advanced and robust query-answering capability.

## 5. RESULTS

The developed application supports visual analytics by enabling users to answer a range of inference questions through inspection of hard and soft graphs. To illustrate its capabilities, several example scenarios are described below. While context is provided here for clarity, the analysis assumes that the observer relies solely on the graph data, without knowledge of the original scenes.

The graphs are visualized as either hard or soft based on the type of data used to construct them. In these graphs, nodes represent key pieces of information, while edges denote the relationships between them. Both graph types use color-coding for nodes based on their category for better visual analytics—Entity, Group, Location, or Activity—and for edges based on the types of nodes they connect, such as Human-Vehicle, Entity-Group, Entity-Location, Entity-Activity, or Activity-Activity. This visual representation allows intuitive interpretation of the situation by examining the nodes and their interconnections.

Example-1 (Package Drop and Pickup scenario): In this scenario, a person enters a room carrying a box, sets it down, leaves, and shortly afterward another person picks it up and exits. Figure 4 shows the hard graph while Figure 5 shows the soft graph for Example-1.

- Hard graph: The sequence is clear. Human H-1 carries object O-1 at 1:23 PM, drops it, and departs. Human H-2 picks up O-1 shortly thereafter. All entities are connected through their shared location: the classroom interior.

- Soft graph: Provides richer entity descriptions—H-1 is a woman in green, H-2 a woman in black, and they belong to separate groups, linked only by their interaction with the object (a cardboard box). While entity details are enhanced, temporal order is obscured. It is only stated that the woman in green dropped the box and the woman in black picked it up, without clear timing. Inference is required to reconstruct the chain of events.

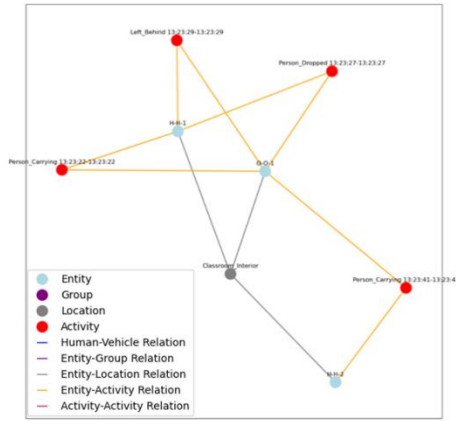This pattern—temporal clarity in hard graphs versus descriptive richness in soft graphs—emerges consistently.
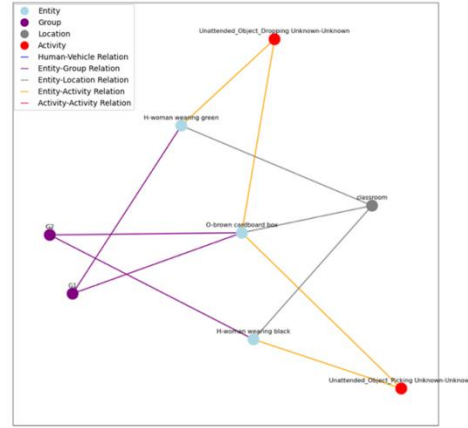
Figure 4: Hard Data Graph for Example 1               Figure 5: Soft Data Graph for Example 1

**Example- 2 (Attempted Delivery Rejected):** A man and his son attempt to deliver items to a woman who refuses to let them inside. The man eventually leaves and discards one item by throwing it onto a garage roof. The hard graph and soft graph representation for example-2 is shown below in Figure 6 and 7 respectively.

- Hard graph: Entities form two groups: G1 (H-1, H-2, O-1, O-2, V-2) and G2 (H-2, H-3, V-1), with H-2 appearing in both, suggesting movement between groups. Simultaneous distant interactions imply brief communication. Vehicle and human associations imply usage or ownership. The sequence includes trunk access, object handoff between individuals, and abandonment of one object.

- Soft graph: Entities are grouped together and described as an older man, a younger man, and a woman. The object set includes a backpack, a package, and pizza—more than the two objects in the hard graph. The location is described as a house. Combining both graphs, it can be inferred that H-1 is the older man, H-2 the younger man, and H-3 the woman, and that the intention was to deliver the objects—and possibly H-2—to the woman's residence.
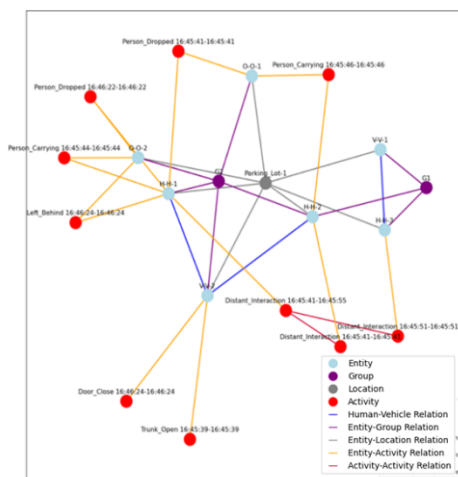


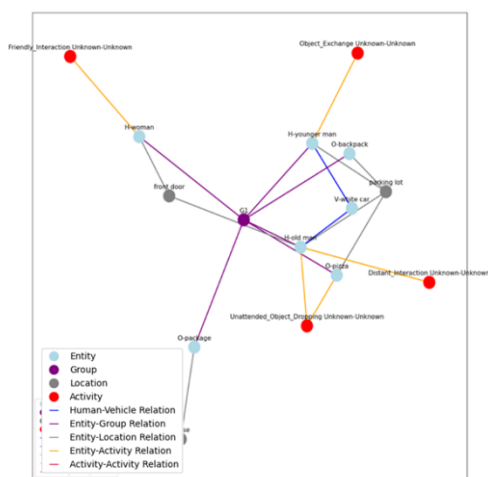Figure 6: Hard Data Graph for Example               Figure 7: Soft Data Graph of Example 2

Example-3 (UPS Delivery): A delivery person pushes a box across an icy driveway, after which the homeowner immediately retrieves it. Example-3's hard graph and soft graph representation is demonstrated in Figure 8 and 9 respectively.

- Hard graph: Events are sequential and clear. Human H-1 carries object O-1, drops it, and H-2 picks it up soon after. H-1, O-1, and vehicle V-1 are part of the same group; H-2 is not. The entire interaction occurs in a parking lot.
- Soft graph: Adds detail about entities and locations. H-1 is the UPS driver, V-1 is the UPS truck, O-1 is the box, and H-2 is the homeowner. Entity locations include driveway, doorway, and house. However, the soft graph misattributes an action: it indicates the UPS truck is "loading" an object, an event that did not occur. Without hard graph support, this could lead to a misinterpretation—that the homeowner handed something to the driver, who then loaded it into the truck. When combined with the hard graph, the correct interpretation is restored: the driver delivered the box to the homeowner, and no loading action took place. The erroneous soft label likely resulted from a misclassification during data processing.

While many questions can currently be answered through graph inspection, others—such as those involving message confidence—require further development. Once hard and soft graphs are fused, confidence metrics can help resolve discrepancies like the soft graph's misclassified loading in Example-3.
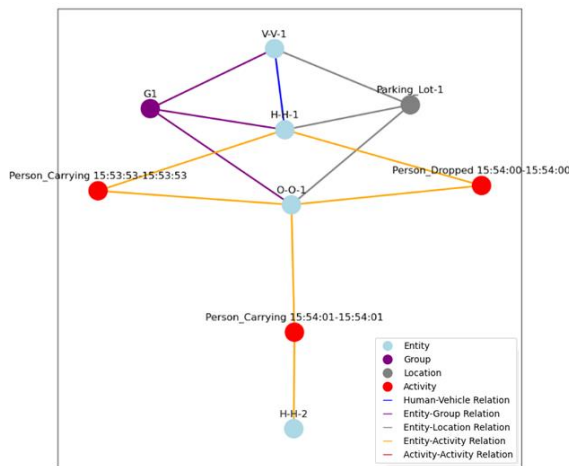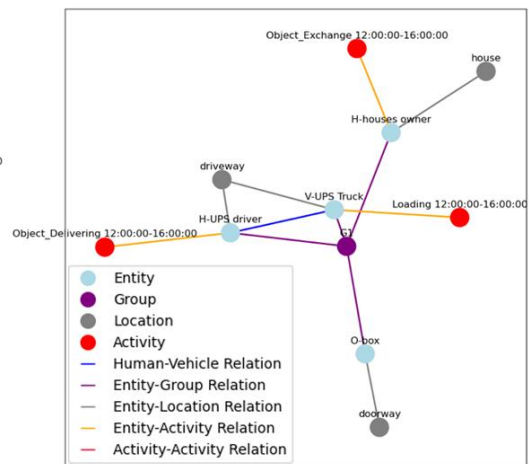


Figure 8: Hard Data Graph for Example 3          Figure 9: Soft Data Graph of Example 3

## 6. CONCLUSION

This work extended the ChatIE algorithm developed by X. Wei et al. [11] to create a systematic methodology for converting unstructured text into structured Transducer Markup Language (TML) messages. With further refinement, this method can be generalized for transforming unstructured input into various structured formats, making it a valuable tool for future research in information extraction and fusion. A foundational implementation of graph-based data fusion was also developed, enabling visual analytics through separate hard and soft graphs, alongside a preliminary automatic query system. Together, these components establish a strong platform for advancing the broader task of hard-soft data fusion.

Despite these accomplishments, several challenges remain. The modified ChatIE pipeline occasionally combines multiple events into a single message, resulting in downstream misclassifications that complicate data fusion. Additionally, the current TML schema lacks sufficient granularity in attribute definitions, limiting classification accuracy when encountering novel events. A further concern is ChatGPT's closed-source nature; changes to its internal behavior could unpredictably affect extraction quality. Future work should explore open-source alternatives that maintain similar accuracy and flexibility. The graph-based fusion system can also be enhanced. Most critically, the fusion of hard and soft graphs via shared entity attributes—the core goal of this approach—remains incomplete. Another promising direction is to reverse the process by using structured hard data to generate soft data messages. Applying natural language processing techniques like document similarity on these generated texts could offer a novel pathway to data fusion. The automatic query system, though functional, requires improvements in its ability to incorporate message confidence and support more complex queries. Such capabilities are essential for scaling inference across large datasets. Finally, a comprehensive comparison of fusion techniques—including Bayesian and Evidence Fusion—remains an open research area. Implementing all three methods on a common dataset would require the development of appropriate performance metrics, as no standard currently exists. Moreover, the lack of publicly available datasets for hard-soft fusion is a significant barrier. Publishing an open-source benchmark dataset would be a valuable contribution, enabling rigorous comparisons and supporting future innovation in this field.

## REFERENCES

[1]     G. A. Gross, R. Nagi, K. Sambhoos, D. R. Schlegel, S. C. Shapiro et al., "Towards hard+soft data fusion: Processing architecture and implementation for the joint fusion and analysis of hard and soft intelligence data," IEEE Conference Publication | IEEE Xplore, Jul. 01, 2012.

[2]     K. Date, G. A. Gross, and R. Nagi, "Test and evaluation of data association algorithms in hard+soft data fusion," IEEE Conference Publication | IEEE Xplore, Jul. 01, 2014.

[3]     J. Llinas, R. Nagi, D. Hall, and J. Lavery, "A Multi-Disciplinary University Research Initiative in Hard and Soft information fusion: Overview, research strategies and initial results," IEEE Conference Publication | IEEE Xplore, Jul. 01, 2010.

[4]     J. R. Chapman, J. L. Crassidis, D. Kasmier, D. Limbaugh, S. Gagnon et al., "Conceptual spaces for space event characterization via hard and soft data fusion," AIAA Scitech 2021 Forum, Jan. 2021, doi: 10.2514/6.2021-1163.

[5]     N. R. Ahmed, E. M. Sample, and M. Campbell, "Bayesian Multicategorical soft data fusion for Human–Robot Collaboration," IEEE Journals & Magazine | IEEE Xplore, Feb. 01, 2013.

[6]     S. Reece, S. Roberts, D. Nicholson, and C. Lloyd, "Determining intent using hard/soft data and Gaussian process classifiers," IEEE Conference Publication | IEEE Xplore, Jul. 01, 2011.

[7]     K. Vasnier, A. Mouaddib, S. Gatepaille, and S. Brunessaux, "Multi-Level Information Fusion Approach with Dynamic Bayesian Networks for an Active Perception of the environment," IEEE Conference Publication | IEEE Xplore, Jul. 01, 2018.

[8]     J. R. Chapman, D. Kasmier, J. L. Crassidis, J. Llinas, B. Smith et al., "Implementing Dampster-Shafer theory for property similarity in conceptual spaces modeling," AIAA SCITECH 2022 Forum, Jan. 2022, doi: 10.2514/6.2022-1272.

[9]     T. L. Wickramarathne, K. Premaratne, M. N. Murthi, M. Scheutz, S. Kübler et al., "Belief theoretic methods for soft and hard data fusion," IEEE Conference Publication | IEEE Xplore, May 01, 2011.

[10]    S. Acharya and M. Kam, "Evidence combination for hard and soft sensor data fusion," IEEE Conference Publication | IEEE Xplore, Jul. 01, 2011.

[11]    X. Wei, X. Cui, N. Cheng, X. Wang, X. Zhang et al., "ChatIE: Zero-Shot Information Extraction via Chatting with ChatGPT," arXiv.org, Feb. 20, 2023.

[12]    V. Elangovan, A. Alkilani and A. Shirkhodaie, "A Multi-Modality Attributes Representation Scheme for Group Activity Characterization and Data Fusion", ISI, pp. 85-90, 2013.

[13]    R. Han, T. Peng, C. Yang, B. Wang, L. Liu et al, "Is Information Extraction Solved by ChatGPT? An Analysis of Performance, Evaluation Criteria, Robustness and Errors", arXiv.org, Sep. 10, 2024.

**AUTHORS**

**Nicholas Gahman** received his B.S. in Computer Science and his Masters in Artificial Intelligence from Penn State University in 2024. His research interests broadly encompass artificial intelligence, with a specific focus on natural language processing. This paper stems from his work at Penn State University.

**Vinayak Elangovan** is an Associate Professor and Program Chair of Computer Science at Penn State Abington. With over 15 years of teaching and research experience, his expertise spans artificial intelligence, machine vision, sensor data fusion, robotics, and activity sequence analysis. His research addresses applications in battlefield intelligence, homeland security, healthcare, manufacturing, and technologies supporting elderly assistance