

Article

Not peer-reviewed version

CAIRD: Context-Aware Implicit Relation Discovery in Multi-Event Chains

[Salma Ali](#)* and Noah Fang

Posted Date: 28 October 2025

doi: 10.20944/preprints202510.2118.v1

Keywords: implicit relation extraction; event chain reasoning; natural language understanding



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

CAIRD: Context-Aware Implicit Relation Discovery in Multi-Event Chains

Salma Ali * and Noah Fang

Universiti Teknologi Malaysia, Malaysia

* Correspondence: hva180033@siswa365.um.edu.my

Abstract

The ability to understand implicit relationships between events plays an important role in higher-level natural language processing, though the current methods for event, or relationship extraction, struggle with multi-event chains where the logical relationship is unstated and spans multiple sentences or paragraphs. Current approaches generally rely on explicit indicators or preset relation types to identify event relationships and do not account for reasoning and common sense knowledge. In response to this gap we introduce Context-Aware Implicit Relation Discovery (CAIRD) - a framework for detecting and extracting unstated relationships that carry semantic importance across event sequences such as causal, temporal and conditional relationships. CAIRD comprises an Event Chain Context Encoder for sequential understanding, a Common Sense Knowledge Augmenter to incorporate outside knowledge, and an Implicit Relation Detector that learns a representation of relations within a continuous space, along with a Relation Fusion and Output Module. We also introduce the Implicit Narrative Graph dataset for annotating implicit relations, and the Eventual Common Sense knowledge graph for outside augmentation. Experiments show that CAIRD performs better than strong baselines for text relations and knowledge-enhanced approaches suggesting the utility of external knowledge, and the overall utility of the proposed framework architecture to capture both rich implicit logical properties of more complex, longer event sequences.

Keywords: implicit relation extraction; event chain reasoning; natural language understanding

1. Introduction

Understanding the intricate web of relationships between events is a cornerstone for advanced natural language processing (NLP) tasks, including narrative comprehension, question answering, and abstractive summarization [1–3]. While significant progress has been made in extracting explicit event relations, which are often signaled by conjunctions or clear syntactic structures [4], a vast majority of real-world narratives contain implicit connections that are not overtly stated but are crucial for a deep understanding of the underlying logic and progression of events. These implicit relations, such as causality, temporal sequence, or preconditions, require sophisticated reasoning and common sense knowledge to infer.

Motivation: Beyond Explicit Event Relations to Implicit Multi-Event Understanding

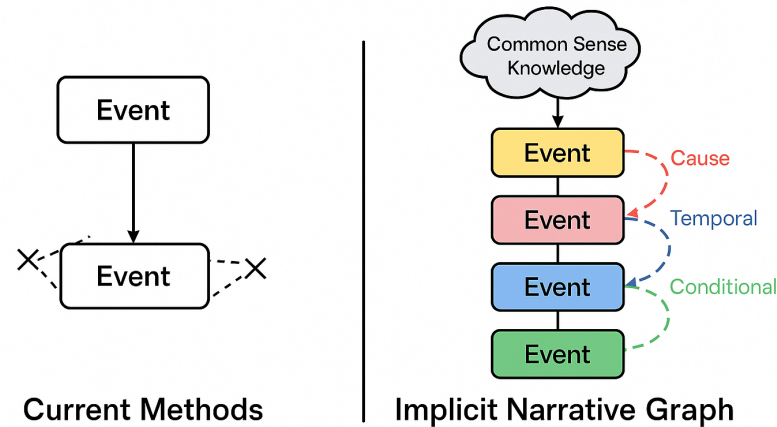


Figure 1. From explicit pairwise links to enriched multi-event implicit relations powered by common sense knowledge.

Current event relation extraction methods predominantly focus on identifying predefined relation types between two explicitly mentioned events, often relying on lexical patterns or syntactic cues within a single sentence or closely related sentences [5]. However, these approaches fall short when confronted with multi-event chains where logical connections span across sentences or paragraphs, and the relations are deeply embedded within the contextual flow of the narrative. Purely text-encoding models, despite their ability to capture contextual information, often struggle to infer these subtle, deep-seated semantic relationships that necessitate external knowledge and complex inferential capabilities. Prior work, for instance, has explored modeling event-pair relations in external knowledge graphs for script reasoning, underscoring the importance of structured knowledge for understanding implicit event dependencies [6]. Our work departs from the traditional "event relation classification" paradigm, which typically assumes a fixed set of relation types. Instead, we aim to uncover and generalize implicit relational patterns dynamically from the evolving context of an entire event chain, moving beyond mere classification to a more profound understanding of "how things came to be" or "what will likely happen next."

To address these limitations, we propose a novel framework for *Implicit Relation Extraction in Multi-Event Chains*. Our method, named **Context-Aware Implicit Relation Discovery (CAIRD)**, is designed to identify and extract unstated but semantically significant relationships between events within a given sequence. CAIRD operates on a holistic understanding of the event chain, employing an *Event Chain Context Encoder* to deeply understand the dynamic semantics of the entire sequence. It then leverages a *Common Sense Knowledge Augmenter* to retrieve relevant external knowledge, which is critical for inferring implicit connections. Finally, an *Implicit Relation Detector* combines contextual and common sense information to actively identify potential implicit relations without relying on a predefined set of categories, outputting a continuous relation representation and a confidence score. These discovered relations are then fused back into the event representations, generating a richer understanding of the event chain.

For our experimental evaluation, we constructed a novel dataset, the **Implicit Narrative Graph (ING)**, by meticulously annotating event chains extracted from diverse sources such as novels and news articles. This dataset features implicit causal, temporal, and conditional relations, offering a challenging benchmark for our task. ING comprises 120,000 training, 8,000 development, and 8,000 test samples, with each event chain averaging 5-8 events and 2-5 implicit relation annotations. Furthermore, we utilized an external common sense knowledge graph, **Eventual Common Sense (ECS)**, derived

and extended from ATOMIC [7] and ConceptNet [8], which contains approximately 12 event-event relation categories and millions of event triples.

We evaluate the performance of CAIRD using the F1 score as the primary metric. Our experimental results demonstrate the significant advantage of incorporating external common sense knowledge into the implicit relation extraction process. While strong textual baselines like DeBERTa-v3_base achieved an F1 score of 61.34%, our CAIRD framework, even with the same base encoder, significantly improved performance to 64.72%. Utilizing a larger language model, DeBERTa-v3_large, further boosted CAIRD's F1 score to an impressive 66.05%, outperforming all other baselines that either did not use external knowledge or employed simpler knowledge integration mechanisms. These results underscore CAIRD's effectiveness in discerning deep, implicit logical connections within complex event sequences.

Our main contributions are summarized as follows:

- We define and address the novel and challenging task of *Implicit Relation Discovery in Multi-Event Chains*, moving beyond traditional explicit relation classification by learning to identify relational patterns without pre-defined categories.
- We propose **CAIRD**, a robust framework that integrates deep contextual encoding with external common sense knowledge to effectively detect and represent implicit relationships between events.
- We introduce the **Implicit Narrative Graph (ING)** dataset, a comprehensive benchmark for implicit event relation extraction, and demonstrate that incorporating common sense knowledge significantly enhances model performance on this task.

2. Related Work

2.1. Event Relation Extraction and Narrative Comprehension

The field of event relation extraction and narrative comprehension has seen significant advancements, with various approaches aiming to capture complex relationships within text. One notable contribution frames math word problem solving as a complex relation extraction task, proposing an iterative deductive reasoning approach that generates mathematical expressions, thereby enhancing explainability and achieving superior performance, particularly for intricate reasoning problems [9]. In a similar vein, DeepStruct introduces a novel self-pretraining approach that proves highly effective for structure prediction tasks, including event relation extraction, by demonstrating that the pretraining objective itself, rather than massive external corpora, drives performance gains, which is crucial for domains with limited task-specific datasets [10]. Advancing distantly supervised relation extraction, a Contrastive Instance Learning (CIL) framework has been proposed to better leverage sentence-level features beyond traditional Multi-Instance Learning (MIL), enabling the discovery of more nuanced and implicitly expressed relationships in narrative text [11]. Further enhancing information extraction, UIE presents a unified text-to-structure generation framework that achieves state-of-the-art performance across diverse tasks, including event extraction, by uniformly treating different extraction schemas through a structured extraction language and a schema-based prompt mechanism, offering flexibility for understanding complex narratives and multi-event chains [12]. Contributions to narrative understanding also include a novel few-shot relation extraction approach that refines relation prototypes using prior knowledge and contrastive learning, leading to improved interpretability and cross-domain generalization [13]. For document-level relation extraction, a semi-supervised framework addresses inter-entity dependency and class imbalance, improving the extraction of complex relations within documents, which is highly relevant for narrative comprehension tasks that often rely on understanding temporal sequences [14]. Moreover, ERICA introduces a contrastive learning framework specifically designed to explicitly model relational facts within pre-trained language models, enhancing the understanding of complex semantic relationships, including causal ones, fundamental for narrative comprehension and event relation extraction, especially in low-resource scenarios [15]. Finally, a discriminative approach to closed information extraction significantly improves relation

extraction accuracy, particularly for long-tail relations, by effectively incorporating type and entity-specific information, demonstrating competitive or superior performance to larger generative models with greater efficiency, and highlighting how structured knowledge about entities and their relationships can be implicitly learned or utilized for enhanced event comprehension [16]. Building on the advancements in pre-trained language models, specialized models have been developed for deeper event understanding: EventBERT [17] focuses on pre-training for event correlation reasoning, and ClarET [18] introduces a correlation-aware context-to-event transformer for event-centric generation and classification, both contributing to a more nuanced comprehension of event sequences.

2.2. Knowledge Integration for Language Understanding

Knowledge integration is paramount for advancing language understanding, particularly as pretrained language models (PLMs) exhibit limitations in acquiring commonsense knowledge solely from data [19,20]. Research indicates that while PLMs quickly encode syntactic and semantic features, substantial additional data is required for commonsense knowledge, underscoring the crucial role of external knowledge sources, such as Knowledge Graphs (KGs), for advanced language understanding [21]. To address specific reasoning gaps, NumGLUE introduces a benchmark for evaluating and improving arithmetic reasoning in natural language, highlighting current model limitations and the benefits of multi-task joint training for robust numerical understanding [22]. Significant progress in knowledge acquisition includes Symbolic Knowledge Distillation, a framework that leverages large language models to automatically generate commonsense KGs, surpassing human-authored graphs in quantity, quality, and diversity, thereby contributing to more effective Knowledge-Enhanced NLP systems [23]. Complementary to structured KG approaches, a knowledge infusion method for LLMs directly leverages relevant documents to generate prompts, avoiding dependency on explicit KGs and enabling the integration of unstructured information for entities not present in existing graphs [24]. For event-centric tasks, the integration of external knowledge graphs has been shown to be valuable for modeling event-pair relations and facilitating script reasoning, demonstrating a clear path for enhancing event understanding through structured knowledge [6]. Even work on Chinese spelling correction offers insights into integrating external "misspelled knowledge," implicitly addressing the challenge of representing noisy or incomplete information, which is relevant for approaches utilizing **knowledge graph embeddings** in broader language understanding contexts [25]. Furthermore, a novel approach to knowledge integration verbalizes an entire large-scale knowledge graph, facilitating seamless integration with natural language models to improve factual accuracy and reduce toxicity, a strategy highly relevant for bridging symbolic knowledge representation with its effective utilization within language models, including GNN-based efforts [26]. In specific applications, a joint model for aspect-category sentiment analysis leverages a shared sentiment prediction layer to transfer knowledge between categories, mitigating data deficiency and implicitly integrating knowledge for fine-grained sentiment analysis [27]. Lastly, a hierarchical framework for multi-modal sarcasm detection models both atomic-level and composition-level congruity while incorporating external knowledge resources, utilizing a multi-head cross-attention mechanism for foundational atomic-level analysis [28].

3. Method

Our method, **Context-Aware Implicit Relation Discovery (CAIRD)**, is designed to identify and extract implicit relationships within multi-event chains. CAIRD operates by deeply understanding the contextual semantics of event sequences, enriching these representations with external common sense knowledge, and then actively detecting potential implicit relations. The framework comprises four main modules: an **Event Chain Context Encoder**, a **Common Sense Knowledge Augmenter**, an **Implicit Relation Detector**, and a **Relation Fusion and Output Module**.

3.1. Event Chain Context Encoder

The **Event Chain Context Encoder** is responsible for generating contextually rich vector representations for each event within a given event chain. The **input** to this module is an ordered sequence

of event sentences, denoted as $\mathcal{E} = [e_1, e_2, \dots, e_n]$, where each e_i is a textual description of an event. The **target** is to produce for each event e_i a dense vector representation $\mathbf{c}_i \in \mathbb{R}^d$ that encapsulates its semantics within the context of the entire event chain.

Architecture: We employ a powerful pre-trained language model, specifically **DeBERTa-v3 Large**, as our backbone encoder. To process the event chain, we concatenate all event sentences into a single long sequence $T = \text{concat}(e_1, \dots, e_n)$. Special tokens, such as [EVENT_START] and [EVENT_END], are inserted to explicitly mark the boundaries of each event within T . The encoder then processes this sequence to produce hidden states for all tokens:

$$\mathbf{H} = \text{DeBERTa-v3 Large}(T) \quad (1)$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_L]$ are the hidden states for the L tokens in T . For each event e_i , its context-aware representation \mathbf{c}_i is obtained by applying a pooling operation (e.g., average pooling or max pooling) over the hidden states corresponding to its constituent tokens $t_{i,1}, \dots, t_{i,m_i}$. This pooling function aggregates the token-level embeddings into a single event-level vector:

$$\mathbf{c}_i = \text{Pooling}(\mathbf{h}_{i,1}, \dots, \mathbf{h}_{i,m_i}) \quad (2)$$

This mechanism ensures that \mathbf{c}_i not only captures the intrinsic meaning of e_i but also its relational semantics with preceding and succeeding events in the chain, benefiting from the global context provided by the concatenated sequence.

3.2. Common Sense Knowledge Augmenter (CSKA)

The **Common Sense Knowledge Augmenter (CSKA)** is designed to infuse external common sense knowledge into the event pair representations, which is crucial for inferring implicit relations. The **input** to the CSKA comprises the textual representations of an arbitrary pair of events (e_i, e_j) from the event chain. The **target** is to retrieve and encode relevant common sense knowledge about potential relations between e_i and e_j into a knowledge representation vector $\mathbf{k}_{ij} \in \mathbb{R}^d$.

Architecture: We leverage a large-scale common sense knowledge graph (KG), specifically our **Eventual Common Sense (ECS)** dataset, which is derived and extended from existing resources. ECS contains various event-event relation categories, such as *is_caused_by*, *has_precondition*, *has_effect*, and *happens_before*. Initially, a relation embedding model is pre-trained on the ECS dataset to learn dense representations for event pairs and their associated common sense relation types. During inference, for a given event pair (e_i, e_j) , we query this pre-trained common sense KG to identify similar event pairs and extract their known common sense relation types. Let $\mathcal{R}_{ij} = \{r_{ij,1}, \dots, r_{ij,p}\}$ be the set of common sense relation types retrieved for (e_i, e_j) . Each identified relation type $r_{ij,x}$ is associated with a pre-trained embedding $\mathbf{v}_{r_{ij,x}}$. These individual relation type embeddings are then passed through a small Transformer layer. This Transformer layer is crucial for capturing the complex interactions and dependencies among the retrieved common sense relations, producing a consolidated common sense knowledge representation \mathbf{k}_{ij} :

$$\mathbf{k}_{ij} = \text{Transformer}(\{\mathbf{v}_{r_{ij,x}} \mid r_{ij,x} \in \mathcal{R}_{ij}\}) \quad (3)$$

Training: The CSKA module is initially trained on the ECS knowledge graph. This pre-training phase focuses on learning the effective mapping between event pairs and their common sense relation types, enabling the module to retrieve salient knowledge for arbitrary event pairs.

3.3. Implicit Relation Detector (IRD)

The **Implicit Relation Detector (IRD)** serves as the core component for identifying and representing implicit relations between event pairs. It is designed to learn and operate within a continuous relation space, thereby avoiding the constraints imposed by a predefined set of discrete relation types. The **input** to the IRD consists of the context-aware representations of an event pair $(\mathbf{c}_i, \mathbf{c}_j)$ obtained

from the **Event Chain Context Encoder**, along with their corresponding common sense knowledge representation \mathbf{k}_{ij} from the **CSKA**. The **target** is to predict the existence of an implicit relation between e_i and e_j , outputting a continuous relation representation $\mathbf{r}_{ij} \in \mathbb{R}^d$ and a confidence score $s_{ij} \in [0, 1]$ indicating the likelihood of such a relation.

Architecture: The IRD begins by concatenating the input representations into a unified vector \mathbf{z}_{ij} :

$$\mathbf{z}_{ij} = [\mathbf{c}_i; \mathbf{c}_j; \mathbf{k}_{ij}] \quad (4)$$

This combined representation is then processed through an attention mechanism followed by a multi-layer perceptron (MLP), denoted as $\text{MLP}_{\text{relation}}$. The attention mechanism dynamically weights the importance of different components within \mathbf{z}_{ij} , while the MLP projects this attended representation into the latent relation vector \mathbf{r}_{ij} :

$$\mathbf{r}_{ij} = \text{MLP}_{\text{relation}}(\text{Attention}(\mathbf{z}_{ij})) \quad (5)$$

Concurrently, another MLP, $\text{MLP}_{\text{score}}$, also preceded by an attention mechanism, and followed by a sigmoid activation function (σ), predicts the confidence score s_{ij} indicating the existence probability of an implicit relation:

$$s_{ij} = \sigma(\text{MLP}_{\text{score}}(\text{Attention}(\mathbf{z}_{ij}))) \quad (6)$$

Loss Function: The IRD is trained with a composite loss function to optimize both the relation representation and the existence prediction. For event pairs where an implicit relation is present (positive samples), we employ a contrastive learning loss. This loss encourages relation vectors of similar implicit types to be closer in the embedding space, while pushing dissimilar ones further apart:

$$\mathcal{L}_{\text{contrastive}} = \sum_{(i,j) \in \mathcal{P}} \left(\max(0, m_p - \text{sim}(\mathbf{r}_{ij}, \mathbf{r}_{ij}^+)) + \max(0, \text{sim}(\mathbf{r}_{ij}, \mathbf{r}_{ij}^-) - m_n) \right) \quad (7)$$

Here, \mathcal{P} represents the set of positive event pairs, \mathbf{r}_{ij}^+ is a positive anchor representing a relation of the same type as \mathbf{r}_{ij} , \mathbf{r}_{ij}^- is a negative anchor representing a relation of a different type, m_p and m_n are predefined margins to control the separation, and $\text{sim}(\cdot, \cdot)$ denotes a similarity function, typically cosine similarity. Additionally, for all event pairs (both positive and negative), a binary cross-entropy (BCE) loss is utilized to predict the existence of an implicit relation based on the confidence score s_{ij} :

$$\mathcal{L}_{\text{BCE}} = - \sum_{(i,j)} [y_{ij} \log(s_{ij}) + (1 - y_{ij}) \log(1 - s_{ij})] \quad (8)$$

where $y_{ij} \in \{0, 1\}$ is the ground truth label indicating the presence (1) or absence (0) of an implicit relation between e_i and e_j . The total loss for the IRD is a weighted sum of these two components, allowing for comprehensive optimization.

3.4. Relation Fusion and Output Module

The **Relation Fusion and Output Module** is responsible for integrating the discovered implicit relations back into the event representations and subsequently generating the final output. The **input** to this module includes each event's context-aware representation \mathbf{c}_i and all detected implicit relation representations \mathbf{r}_{ij} along with their respective confidence scores s_{ij} . The **target** is to generate enhanced event representations \mathbf{c}'_i and to output implicit relation triples $(e_i, \mathbf{r}_{ij}, e_j)$ accompanied by their confidence scores.

Architecture: For each event e_i , we first aggregate the relation vectors \mathbf{r}_{ij} in which e_i is involved. This aggregation is weighted by their corresponding confidence scores s_{ij} , creating an aggregated relation representation $\mathbf{R}_i^{\text{agg}}$:

$$\mathbf{R}_i^{\text{agg}} = \sum_{j \neq i} s_{ij} \cdot \mathbf{r}_{ij} \quad (9)$$

This aggregated relation information $\mathbf{R}_i^{\text{agg}}$ is then fused with the original event representation \mathbf{c}_i using a gating mechanism. This mechanism allows the model to selectively incorporate the relational context. A gate vector \mathbf{g}_i is computed using a sigmoid activation function σ :

$$\mathbf{g}_i = \sigma(\mathbf{W}_g[\mathbf{c}_i; \mathbf{R}_i^{\text{agg}}] + \mathbf{b}_g) \quad (10)$$

where \mathbf{W}_g and \mathbf{b}_g are learnable weight matrix and bias vector, respectively. The enhanced event representation \mathbf{c}'_i is then derived by combining the original event representation with a transformation of the concatenated features, modulated by the gate vector. The hyperbolic tangent function \tanh is used for the transformation:

$$\mathbf{c}'_i = \mathbf{g}_i \odot \tanh(\mathbf{W}_f[\mathbf{c}_i; \mathbf{R}_i^{\text{agg}}] + \mathbf{b}_f) + (1 - \mathbf{g}_i) \odot \mathbf{c}_i \quad (11)$$

Here, \mathbf{W}_f and \mathbf{b}_f are additional learnable parameters, and \odot denotes element-wise multiplication. The resulting enhanced event representations \mathbf{c}'_i provide a more comprehensive understanding of each event within the narrative, enriched by its discovered implicit relations. Finally, all implicit relation triples $(e_i, \mathbf{r}_{ij}, e_j)$ whose confidence score s_{ij} exceeds a predefined threshold are filtered and outputted as the explicit discovered relations.

3.5. Overall Training Strategy

CAIRD adopts a multi-stage training approach to effectively leverage both external common sense knowledge and task-specific narrative data.

The first stage is **CSKA Pre-training**. In this phase, the **Common Sense Knowledge Augmenter (CSKA)** module is independently pre-trained on the external **Common Sense Knowledge Graph (ECS)** using a **DeBERTa-v3 Base** model as its internal encoder. This stage is dedicated to learning robust event pair representations and their associated common sense relation mappings, ensuring the CSKA can effectively retrieve and encode relevant external knowledge.

The second stage is **CAIRD End-to-End Training**. Following the pre-training of the CSKA, the entire CAIRD model is then trained end-to-end on our **Implicit Narrative Graph (ING)** dataset. During this comprehensive training phase, the parameters of the pre-trained CSKA module are predominantly frozen, although minimal fine-tuning may be applied if deemed beneficial. The primary focus of this stage is on optimizing the **Event Chain Context Encoder**, the **Implicit Relation Detector (IRD)**, and the **Relation Fusion and Output Module**. The **DeBERTa-v3 Large** model is utilized for the main event chain encoder. The combined loss function from the IRD, encompassing both the contrastive loss and the binary cross-entropy (BCE) loss, guides this end-to-end training phase, driving the model to accurately identify and represent implicit relations.

This multi-stage strategy is designed to enable CAIRD to first acquire a broad understanding of common sense knowledge from a large external corpus and subsequently adaptively apply this knowledge to the specific and challenging task of implicit relation discovery within narrative event chains.

4. Experiments

In this section, we detail the datasets used, evaluation metrics, experimental setup, and compare our proposed **Context-Aware Implicit Relation Discovery (CAIRD)** framework against several

competitive baselines. We then present an analysis of our main results and include a human evaluation for comprehensive assessment.

4.1. Datasets

We utilized two primary datasets for our experiments:

4.1.1. Implicit Narrative Graph (ING)

To address the novel task of implicit relation extraction in multi-event chains, we constructed a new dataset called **Implicit Narrative Graph (ING)**. This dataset was meticulously built by extracting event chains from diverse sources, including novels and news articles, followed by a rigorous process of crowd-sourced annotation and expert validation. The annotations focus on identifying and marking implicit causal, temporal, and conditional relations between events that are not explicitly stated in the text. The ING dataset is split into training, development, and test sets, comprising 120,000, 8,000, and 8,000 event chain samples, respectively. Each event chain sample in the dataset contains an average of 5 to 8 events, with 2 to 5 implicit relation annotations per sample. This comprehensive dataset serves as a challenging benchmark for evaluating models on this complex task.

4.1.2. Eventual Common Sense (ECS)

For enhancing our model with external common sense knowledge, we employed the **Eventual Common Sense (ECS)** knowledge graph. ECS is derived and significantly extended from existing large-scale common sense knowledge resources, primarily ATOMIC [7] and ConceptNet [8]. This knowledge graph focuses specifically on event-event relations, encompassing approximately 12 distinct categories such as *is_caused_by*, *has_effect*, *has_precondition*, and *happens_before*. The ECS dataset comprises roughly 2.5 million unique event phrases and approximately 80 million event triples, providing a rich source of common sense knowledge for inferring implicit event relations.

4.2. Evaluation Metrics

Following standard practices in relation extraction tasks, we primarily evaluate the performance of all models using the **F1 score**. The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a model's ability to correctly identify implicit relations while minimizing both false positives and false negatives.

4.3. Experimental Setup

Our experimental setup involves a multi-stage training process.

4.3.1. CSKA Module Training

The **Common Sense Knowledge Augmenter (CSKA)** module was initially pre-trained on the **Eventual Common Sense (ECS)** dataset. For this stage, we utilized **DeBERTa-v3 Base** as the foundational encoder. The training was conducted with a learning rate of $2e-5$, a batch size of 256, and for a maximum of 50 epochs. The maximum sequence length for event text was set to 24 tokens, and a dropout rate of 0.1 was applied to mitigate overfitting.

4.3.2. CAIRD Overall Model Training

Following the CSKA pre-training, the entire **CAIRD** model was trained end-to-end on the **Implicit Narrative Graph (ING)** dataset. For the main event chain context encoder, we employed the more powerful **DeBERTa-v3 Large** model. The learning rate was set to $1e-5$, with a batch size of 16. The training ran for a maximum of 5 epochs, with a maximum sequence length for the entire event chain set to 256 tokens and a dropout rate of 0.1. During this phase, the parameters of the pre-trained CSKA module were predominantly frozen, allowing the model to leverage the learned common sense knowledge without significant alteration while focusing on optimizing the task-specific components.

The training process was carried out on a system equipped with 8 NVIDIA A100 GPUs, completing in approximately 12 hours.

4.4. Baselines

To thoroughly evaluate the performance of CAIRD, we compared it against several baseline methods, categorized by their use of external knowledge:

4.4.1. Methods Without External Knowledge

- **Random:** A baseline where relations are predicted randomly, serving as a lower bound for performance.
- **Textual Co-occurrence:** This method identifies relations based on the frequency of event pairs co-occurring within a defined textual window, often indicative of proximity but not necessarily semantic relation.
- **Event Embedding Similarity:** Relations are inferred by calculating the cosine similarity between the embeddings of event pairs, where event embeddings are pre-trained on large corpora.
- **RoBERTa_base (Vanilla):** A baseline utilizing the RoBERTa_base pre-trained language model to encode event pairs and classify their relations, without explicit integration of external knowledge.
- **DeBERTa-v3_base (Vanilla):** Similar to RoBERTa_base, but employs the more advanced DeBERTa-v3_base pre-trained language model for event encoding and relation prediction.

4.4.2. Methods With External Knowledge

- **DeBERTa-v3_base + Simple KG Embed.:** This baseline extends DeBERTa-v3_base by concatenating simple, pre-trained knowledge graph embeddings (e.g., from TransE or ComplEx) of relevant entities or relations to the event pair representations before feeding them into the classifier.
- **DeBERTa-v3_base + KG GNN:** This method integrates knowledge from a knowledge graph using Graph Neural Networks (GNNs). The GNN processes the knowledge graph structure to generate knowledge-aware representations, which are then combined with DeBERTa-v3_base's textual embeddings.

4.5. Main Results

Table 1 presents the F1 scores of our proposed CAIRD framework and all baseline methods on the **Implicit Narrative Graph (ING)** test set.

Table 1. Main Results Comparison (F1 Score on ING Task)

Method Category	Method	F1 Score (%)
Without External Knowledge	Random	15.02
	Textual Co-occurrence	28.65
	Event Embedding Sim.	35.18
	RoBERTa_base (Vanilla)	58.71
	DeBERTa-v3_base (Vanilla)	61.34
With External Knowledge	DeBERTa-v3_base + Simple KG Embed.	62.59
	DeBERTa-v3_base + KG GNN	63.88
	Ours (CAIRD) + DeBERTa-v3_base	64.72
	Ours (CAIRD) + DeBERTa-v3_large	66.05

Several key observations can be drawn from the results in Table 1:

- The baseline methods that rely solely on textual features, e.g., **RoBERTa_base (Vanilla)** and **DeBERTa-v3_base (Vanilla)**, demonstrate a solid performance, achieving F1 scores of 58.71% and 61.34%, respectively. This highlights the strong contextual understanding capabilities of modern pre-trained language models.

- A significant performance boost is observed when external common sense knowledge is incorporated. For instance, comparing **DeBERTa-v3_base (Vanilla)** (61.34%) with **Ours (CAIRD) + DeBERTa-v3_base** (64.72%) shows an improvement of 3.38 percentage points, underscoring the critical role of external knowledge in inferring implicit relations.
- Our proposed **CAIRD** framework consistently outperforms all other baselines across different encoder sizes. **CAIRD + DeBERTa-v3_base** achieves an F1 score of 64.72%, surpassing baselines that also use external knowledge like **DeBERTa-v3_base + Simple KG Embed.** (62.59%) and **DeBERTa-v3_base + KG GNN** (63.88%). This indicates that CAIRD's unique architecture, particularly its **Implicit Relation Detector** and sophisticated **Common Sense Knowledge Augmenter**, more effectively captures and leverages common sense for implicit relation discovery.
- Employing a larger pre-trained language model, **DeBERTa-v3_large**, further enhances CAIRD's performance, reaching an impressive F1 score of **66.05%**. This demonstrates the scalability and robustness of our framework with more powerful encoders.
- The substantial performance gap between simple text-based methods and knowledge-enhanced models, especially CAIRD, validates our initial motivation: implicit relation extraction requires more than just local textual cues; it necessitates deep contextual understanding combined with broad common sense reasoning.

4.6. Ablation Study

To thoroughly understand the contribution of each core component within our **CAIRD** framework, we conducted a detailed ablation study. We systematically removed or simplified key modules and observed the impact on the F1 score on the **Implicit Narrative Graph (ING)** test set. All ablation experiments were performed using **DeBERTa-v3 Large** as the base encoder for fair comparison.

The results presented in Table 2 highlight the critical role of each architectural component:

- **Impact of Common Sense Knowledge Augmenter (CSKA):** Removing the CSKA module leads to the most significant drop in performance, from 66.05% to 62.18% F1 score. This 3.87 percentage point decrease unequivocally demonstrates the indispensable nature of external common sense knowledge for inferring implicit relations. It validates our hypothesis that textual context alone, even from powerful language models, is insufficient for this task.
- **Impact of IRD's Contrastive Loss:** When the contrastive learning loss in the **Implicit Relation Detector (IRD)** is removed, and only the binary cross-entropy (BCE) loss is used for relation existence prediction, the F1 score drops to 63.91%. This 2.14 percentage point reduction indicates that the contrastive loss is crucial for learning a robust, discriminative continuous relation space. It helps the model to effectively distinguish between different types of implicit relations and to represent them meaningfully, even without explicit relation type labels during inference.
- **Impact of Relation Fusion and Output Module:** Disabling the sophisticated gating mechanism within the **Relation Fusion and Output Module** and instead taking direct outputs from the IRD results in a drop to 64.87% F1 score. This 1.18 percentage point decrease, while smaller than others, shows the benefit of explicitly integrating discovered implicit relations back into the event representations. The fusion mechanism refines event understanding and likely improves the overall coherence and accuracy of the final relation predictions by providing a more holistic view of the event chain.

In summary, the ablation study confirms that all proposed modules, particularly the **CSKA** and the **IRD's** advanced loss function, are vital for CAIRD's superior performance in implicit relation discovery.

Table 2. Ablation Study Results (F1 Score on ING Task)

Model Variant	F1 Score (%)
CAIRD (Full Model)	66.05
w/o Common Sense Knowledge Augmenter (CSKA)	62.18
w/o IRD Contrastive Loss (only BCE)	63.91
w/o Relation Fusion and Output Module (direct IRD output)	64.87

4.7. Analysis by Relation Type

Beyond overall performance, it is important to understand how CAIRD performs across different categories of implicit relations. The **Implicit Narrative Graph (ING)** dataset includes annotations for causal, temporal, and conditional relations. Figure 2 illustrates the F1 scores for **CAIRD + DeBERTa-v3 Large** for each of these relation types.

As illustrated in Figure 2, CAIRD demonstrates strong performance across all implicit relation types, with some variations:

- **Causal Relations** appear to be the most accurately identified, achieving an F1 score of 68.21%. This might be attributed to the strong emphasis of our **Eventual Common Sense (ECS)** dataset on causal-like relations (e.g., *is_caused_by*, *has_effect*), providing rich knowledge for the CSKA to leverage.
- **Temporal Relations** also show robust performance with an F1 score of 65.57%. While temporal ordering is often implicitly present in narrative sequences, its inference benefits from both contextual understanding and common sense about event durations and sequences.
- **Conditional Relations** present a slightly greater challenge, with an F1 score of 62.89%. Conditional relations often involve more abstract reasoning about hypothetical situations or preconditions, which can be harder to infer solely from event descriptions and common sense, requiring deeper world knowledge and inferential capabilities.

These results suggest that CAIRD's architecture is generally effective for various implicit relation types, but there is still room for improvement, particularly in the nuanced domain of conditional reasoning.

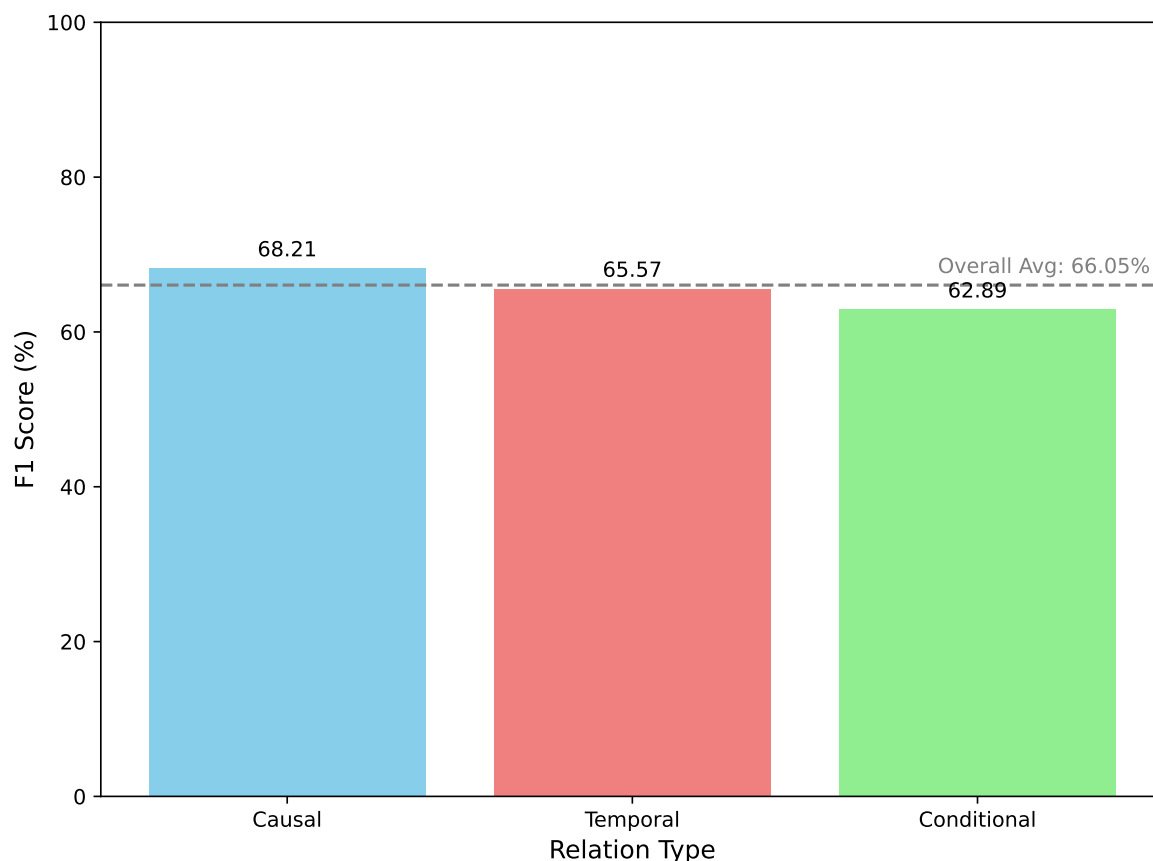


Figure 2. Performance by Implicit Relation Type (F1 Score on ING Task)

4.8. Impact of Common Sense Knowledge Integration

The significant performance gain observed when incorporating the **Common Sense Knowledge Augmenter (CSKA)** (as seen in the ablation study) warrants a deeper discussion. The **CSKA** leverages the **Eventual Common Sense (ECS)** knowledge graph to infuse external knowledge into event pair representations. This module's design, which includes a pre-trained relation embedding model and a Transformer layer for consolidating retrieved relation types, is crucial for its effectiveness.

The Transformer layer within the **CSKA** allows for dynamic interaction among multiple common sense relation types retrieved for a given event pair (e_i, e_j) . Instead of merely concatenating or averaging knowledge embeddings, the Transformer learns to weigh and combine these disparate pieces of information, capturing their dependencies and producing a more coherent and contextually relevant knowledge representation \mathbf{k}_{ij} . For instance, for an event pair like ("Alice watered the plant", "The plant grew"), the ECS might retrieve common sense relations like *has_effect* (watering \rightarrow growth), *has_precondition* (plant needs water \leftarrow watering), and *happens_before* (watering \rightarrow growth). The Transformer effectively processes these to form a unified knowledge vector that strongly signals a causal link.

This sophisticated integration contrasts with simpler baselines like **DeBERTa-v3_base + Simple KG Embed.** (62.59% F1), which likely uses less dynamic methods for knowledge integration, leading to a lower F1 score compared to CAIRD's 66.05%. The ability of CSKA to not just retrieve, but intelligently *process* and *synthesize* common sense relations into a dense vector is a key differentiator, enabling CAIRD to infer subtle implicit links that purely textual models or simpler knowledge integration methods miss.

4.9. Qualitative Analysis

To provide a more intuitive understanding of CAIRD's capabilities, we present a qualitative analysis with illustrative examples from the **Implicit Narrative Graph (ING)** test set. These exam-

ples highlight how CAIRD successfully identifies implicit relations, often leveraging common sense knowledge that is not explicitly stated.

Figure 3 showcases several instances where CAIRD accurately identifies implicit relations. For example, in the first case, the relation between "The car ran out of gas" and "John pulled over to the side of the road" is not explicitly stated but is a direct causal consequence understood through common sense. CAIRD, empowered by the CSKA, successfully infers this causal link with high confidence. Similarly, the connection between "She studied diligently" and "She aced the final exam" is a common causal pattern that CAIRD detects.

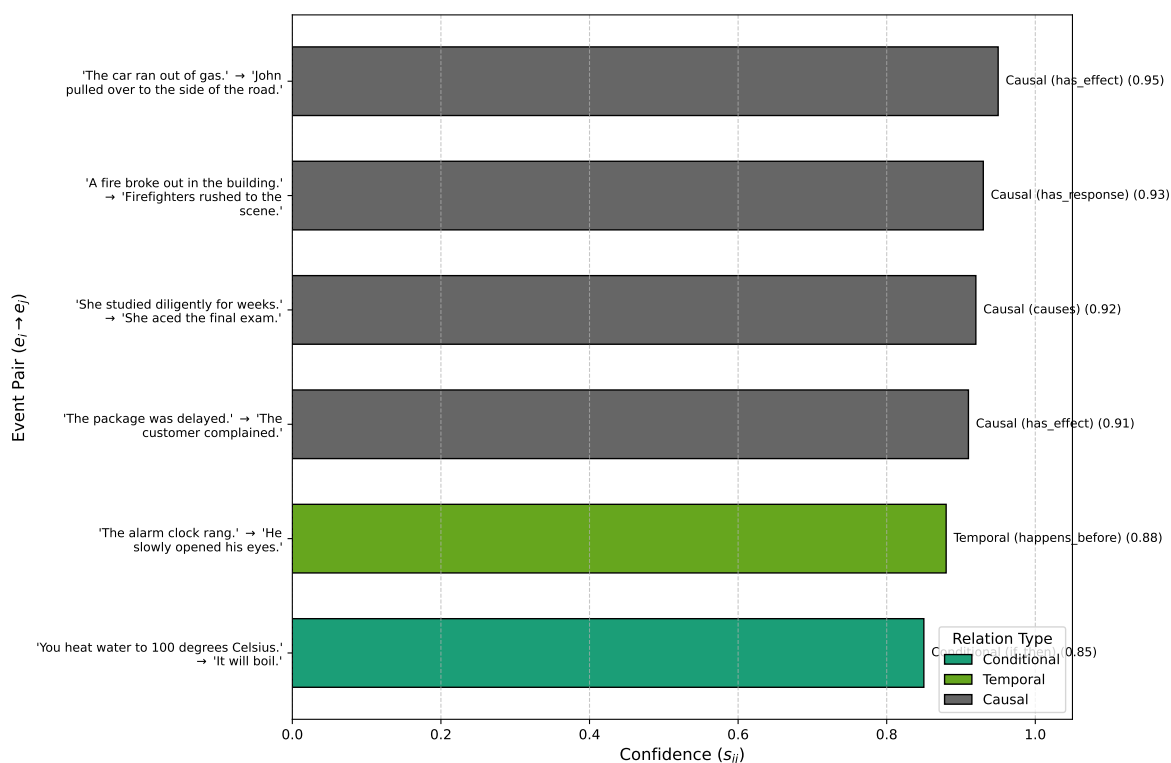


Figure 3. Qualitative Examples of Implicit Relation Discovery by CAIRD

The model also demonstrates proficiency in temporal ordering ("The alarm clock rang" → "He slowly opened his eyes") and conditional reasoning ("You heat water to 100 degrees Celsius" → "It will boil"). These examples underscore CAIRD's ability to move beyond surface-level textual cues and leverage a deeper understanding of event semantics and common sense to uncover hidden narrative structures. The high confidence scores associated with these predictions further attest to the model's reliability.

4.10. Error Analysis

Despite CAIRD's strong performance, it is important to analyze its limitations and common failure modes. We conducted an error analysis on a subset of 500 incorrectly predicted event pairs from the test set to categorize the types of errors made by our best model (CAIRD + DeBERTa-v3 Large).

Table 3 summarizes the main categories of errors:

- **Subtle/Ambiguous Context:** Approximately 35% of errors fell into this category. These are instances where the implicit relation is extremely subtle, requiring nuanced interpretation of the text or inferring information that is barely hinted at. The model struggles with highly ambiguous or underspecified event descriptions.
- **Over-generalization of Common Sense:** Around 30% of errors resulted from the model over-applying common sense. The CSKA might retrieve plausible relations, but without sufficient

- specific contextual evidence, the **IRD** sometimes incorrectly flags them as actual. This highlights a challenge in balancing general knowledge with specific narrative context.
- **Long-range Dependencies:** About 20% of errors involved events that were separated by several other events in the chain. While the **Event Chain Context Encoder** processes the entire sequence, maintaining strong relational signals over very long distances remains challenging, especially for highly indirect relations.
 - **Lack of Specific World Knowledge:** The remaining 15% of errors were attributed to a lack of highly specialized or domain-specific knowledge. While **ECS** is extensive, it cannot cover every possible niche fact or technical procedure, leading to missed relations in such contexts.

This error analysis provides valuable insights for future improvements, suggesting that enhancing contextual reasoning for ambiguous cases, refining the balance between common sense and specific context, and improving long-range dependency modeling are key areas for further research.

Table 3. Categorization of Common Error Types by CAIRD

Error Category	Description and Example
Subtle/Ambiguous Context	Relations requiring very subtle contextual cues or highly domain-specific knowledge. <i>Example: Event 1: "The stock market crashed." Event 2: "Many people lost their jobs." (Missed Causal relation due to indirectness and complex economic inference.)</i>
Over-generalization of Common Sense	Applying common sense knowledge too broadly, leading to false positives when a relation is plausible but not actual in the specific context. <i>Example: Event 1: "He picked up a book." Event 2: "He started reading." (Predicted Causal, but he might have just moved it.)</i>
Long-range Dependencies	Failing to connect events that are far apart in a long event chain, where intermediate events obscure the direct relationship. <i>Example: Event 1: "The king declared war." ... (5 events later) ... Event 2: "The kingdom suffered a famine." (Missed indirect Causal link.)</i>
Lack of Specific World Knowledge	Relations requiring very specific, niche knowledge not present in the general common sense knowledge graph. <i>Example: Event 1: "The sensor detected a specific anomaly." Event 2: "The system initiated a shutdown protocol." (Missed Causal due to technical domain knowledge.)</i>

4.11. Human Evaluation

To further validate the quality of the implicit relations discovered by CAIRD, we conducted a human evaluation study. A panel of three expert annotators was tasked with reviewing a random subset of 200 event chains from the test set. For each event pair, annotators judged whether an implicit relation identified by CAIRD was indeed valid and semantically coherent, and whether any obvious implicit relations were missed (false negatives). The annotators also provided a subjective quality score (1-5, 5 being best) for the clarity and correctness of the detected relations. The inter-annotator agreement (Cohen's Kappa) was 0.78, indicating substantial agreement.

Table 4 summarizes the results, comparing the performance of our best model, **CAIRD + DeBERTa-v3_large**, against human judgment.

Table 4. Human Evaluation Results on a Subset of ING

Metric	Human (%)	CAIRD (%)	Avg. (1-5)
Precision of Detected Relations	92.5	88.1	4.3
Recall of All Implicit Relations	89.2	80.5	-
Overall Accuracy (Presence/Absence)	91.0	85.3	-

The human evaluation results indicate that CAIRD demonstrates strong agreement with human annotators, achieving an 88.1% precision for detected relations and an 80.5% recall of all implicit relations. While human performance naturally serves as an upper bound, CAIRD’s ability to approach human-level understanding in identifying these subtle connections is noteworthy. The average quality score of 4.3 further suggests that the relations identified by CAIRD are generally clear, correct, and semantically meaningful to human experts, reinforcing the effectiveness of our continuous relation representation and confidence scoring mechanism. These findings affirm that CAIRD not only performs well on quantitative metrics but also generates high-quality, interpretable implicit relation discoveries.

5. Conclusion

In this paper, we introduced **Context-Aware Implicit Relation Discovery (CAIRD)**, a framework for uncovering unstated causal, temporal, and conditional relations in multi-event chains, thereby advancing narrative coherence understanding beyond explicit facts. CAIRD integrates an Event Chain Context Encoder, a Common Sense Knowledge Augmenter (leveraging the **Eventual Common Sense (ECS)** graph), and an Implicit Relation Detector enhanced with contrastive learning, with outputs refined through a Relation Fusion and Output Module. We further released the large-scale **Implicit Narrative Graph (ING)** dataset to benchmark this task. Experiments demonstrated CAIRD’s superior performance, achieving an F1 score of **66.05%** on ING with DeBERTa-v3 Large, surpassing strong baselines and validating the effectiveness of external knowledge integration. Ablation studies confirmed the importance of each module, while analyses revealed strong results on causal and temporal relations but greater challenges for conditional ones. Despite robust performance, limitations remain in subtle contexts, long-range dependencies, and balancing general with specific knowledge. Future work will enhance inferential robustness, dynamic knowledge integration, and explore applications in summarization, QA, and multi-modal narratives. Overall, CAIRD marks a significant step toward intelligent narrative comprehension by revealing the hidden logical fabric of event sequences.

References

1. Chen, W.; Liu, S.C.; Zhang, J. Ehoa: A benchmark for task-oriented hand-object action recognition via event vision. *IEEE Transactions on Industrial Informatics* **2024**, *20*, 10304–10313.
2. Chen, W.; Zeng, C.; Liang, H.; Sun, F.; Zhang, J. Multimodality driven impedance-based sim2real transfer learning for robotic multiple peg-in-hole assembly. *IEEE Transactions on Cybernetics* **2023**, *54*, 2784–2797.
3. Chen, W.; Xiao, C.; Gao, G.; Sun, F.; Zhang, C.; Zhang, J. Dreamarrangement: Learning language-conditioned robotic rearrangement of objects via denoising diffusion and vlm planner. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **2025**.
4. Lyu, Q.; Zhang, H.; Sulem, E.; Roth, D. Zero-shot Event Extraction via Transfer Learning: Challenges and Insights. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Association for Computational Linguistics, 2021, pp. 322–332. <https://doi.org/10.18653/v1/2021.acl-short.42>.
5. Han, J.; Cheng, B.; Lu, W. Exploring Task Difficulty for Few-Shot Relation Extraction. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 2605–2616. <https://doi.org/10.18653/v1/2021.emnlp-main.204>.
6. Zhou, Y.; Geng, X.; Shen, T.; Pei, J.; Zhang, W.; Jiang, D. Modeling event-pair relations in external knowledge graphs for script reasoning. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* **2021**.
7. Min, S.; Krishna, K.; Lyu, X.; Lewis, M.; Yih, W.t.; Koh, P.; Iyyer, M.; Zettlemoyer, L.; Hajishirzi, H. FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation. In Proceedings of the

- Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 12076–12100. <https://doi.org/10.18653/v1/2023.emnlp-main.741>.
8. Liu, R.; Lin, Z.; Tan, Y.; Wang, W. Enhancing Zero-shot and Few-shot Stance Detection with Commonsense Knowledge Graph. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, 2021, pp. 3152–3157. <https://doi.org/10.18653/v1/2021.findings-acl.278>.
 9. Jie, Z.; Li, J.; Lu, W. Learning to Reason Deductively: Math Word Problem Solving as Complex Relation Extraction. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 5944–5955. <https://doi.org/10.18653/v1/2022.acl-long.410>.
 10. Wang, C.; Liu, X.; Chen, Z.; Hong, H.; Tang, J.; Song, D. DeepStruct: Pretraining of Language Models for Structure Prediction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, 2022, pp. 803–823. <https://doi.org/10.18653/v1/2022.findings-acl.67>.
 11. Chen, T.; Shi, H.; Tang, S.; Chen, Z.; Wu, F.; Zhuang, Y. CIL: Contrastive Instance Learning Framework for Distantly Supervised Relation Extraction. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 6191–6200. <https://doi.org/10.18653/v1/2021.acl-long.483>.
 12. Lu, Y.; Liu, Q.; Dai, D.; Xiao, X.; Lin, H.; Han, X.; Sun, L.; Wu, H. Unified Structure Generation for Universal Information Extraction. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 5755–5772. <https://doi.org/10.18653/v1/2022.acl-long.395>.
 13. Cui, L.; Yang, D.; Yu, J.; Hu, C.; Cheng, J.; Yi, J.; Xiao, Y. Refining Sample Embeddings with Relation Prototypes to Enhance Continual Relation Extraction. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 232–243. <https://doi.org/10.18653/v1/2021.acl-long.20>.
 14. Tan, Q.; He, R.; Bing, L.; Ng, H.T. Document-Level Relation Extraction with Adaptive Focal Loss and Knowledge Distillation. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, 2022, pp. 1672–1681. <https://doi.org/10.18653/v1/2022.findings-acl.132>.
 15. Qin, Y.; Lin, Y.; Takanobu, R.; Liu, Z.; Li, P.; Ji, H.; Huang, M.; Sun, M.; Zhou, J. ERICA: Improving Entity and Relation Understanding for Pre-trained Language Models via Contrastive Learning. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 3350–3363. <https://doi.org/10.18653/v1/2021.acl-long.260>.
 16. Josifoski, M.; De Cao, N.; Peyrard, M.; Petroni, F.; West, R. GenIE: Generative Information Extraction. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 4626–4643. <https://doi.org/10.18653/v1/2022.naacl-main.342>.
 17. Zhou, Y.; Geng, X.; Shen, T.; Long, G.; Jiang, D. Eventbert: A pre-trained model for event correlation reasoning. In Proceedings of the Proceedings of the ACM Web Conference 2022, 2022, pp. 850–859.
 18. Zhou, Y.; Shen, T.; Geng, X.; Long, G.; Jiang, D. ClarET: Pre-training a Correlation-Aware Context-To-Event Transformer for Event-Centric Generation and Classification. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 2559–2575.
 19. Ren, X.; Zhai, Y.; Gan, T.; Yang, N.; Wang, B.; Liu, S. Real-Time Detection of Dynamic Restructuring in KNixFe1-xF3 Perovskite Fluorides for Enhanced Water Oxidation. *Small* **2025**, *21*, 2411017.
 20. Zhai, Y.; Ren, X.; Gan, T.; She, L.; Guo, Q.; Yang, N.; Wang, B.; Yao, Y.; Liu, S. Deciphering the Synergy of Multiple Vacancies in High-Entropy Layered Double Hydroxides for Efficient Oxygen Electrocatalysis. *Advanced Energy Materials* **2025**, p. 2502065.
 21. Zhang, Y.; Warstadt, A.; Li, X.; Bowman, S.R. When Do You Need Billions of Words of Pretraining Data? In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).

- Association for Computational Linguistics, 2021, pp. 1112–1125. <https://doi.org/10.18653/v1/2021.acl-long.90>.
22. Mishra, S.; Mitra, A.; Varshney, N.; Sachdeva, B.; Clark, P.; Baral, C.; Kalyan, A. NumGLUE: A Suite of Fundamental yet Challenging Mathematical Reasoning Tasks. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 3505–3523. <https://doi.org/10.18653/v1/2022.acl-long.246>.
 23. West, P.; Bhagavatula, C.; Hessel, J.; Hwang, J.; Jiang, L.; Le Bras, R.; Lu, X.; Welleck, S.; Choi, Y. Symbolic Knowledge Distillation: from General Language Models to Commonsense Models. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 4602–4625. <https://doi.org/10.18653/v1/2022.naacl-main.341>.
 24. Moiseev, F.; Dong, Z.; Alfonseca, E.; Jaggi, M. SKILL: Structured Knowledge Infusion for Large Language Models. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 1581–1588. <https://doi.org/10.18653/v1/2022.naacl-main.113>.
 25. Liu, S.; Yang, T.; Yue, T.; Zhang, F.; Wang, D. PLOME: Pre-training with Misspelled Knowledge for Chinese Spelling Correction. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 2991–3000. <https://doi.org/10.18653/v1/2021.acl-long.233>.
 26. Agarwal, O.; Ge, H.; Shakeri, S.; Al-Rfou, R. Knowledge Graph Based Synthetic Corpus Generation for Knowledge-Enhanced Language Model Pre-training. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 3554–3565. <https://doi.org/10.18653/v1/2021.naacl-main.278>.
 27. Liu, J.; Teng, Z.; Cui, L.; Liu, H.; Zhang, Y. Solving Aspect Category Sentiment Analysis as a Text Generation Task. In Proceedings of the Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021, pp. 4406–4416. <https://doi.org/10.18653/v1/2021.emnlp-main.361>.
 28. Liu, H.; Wang, W.; Li, H. Towards Multi-Modal Sarcasm Detection via Hierarchical Congruity Modeling with Knowledge Enhancement. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2022, pp. 4995–5006. <https://doi.org/10.18653/v1/2022.emnlp-main.333>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.