

Article

Enhancing Chinese Event Prediction with Prompt-Driven Knowledge Augmentation

Shulan Lou, Xiaoxue Xie, Wei Liu *  and Wangtao Jiang

School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; loushulan@shu.edu.cn (S.L.); xiexiaoxue@shu.edu.cn (X.X.); jiangwangtao@shu.edu.cn (W.J.)

* Correspondence: liuw@shu.edu.cn

Abstract

Achieving high accuracy in Chinese event prediction remains challenging due to the prevalence of short-text events that lack sufficient background knowledge and semantic depth. Existing approaches predominantly rely on knowledge augmentation strategies that adopt a retrieval-and-integration paradigm, which often results in suboptimal performance due to the limited scope and timeliness of KGs (Knowledge Graphs). Large language models (LLMs) play a crucial role in event knowledge augmentation by serving as dynamic, context-aware knowledge generators that address the limitations of KGs. This paper introduces an enhanced method for Chinese event prediction, named CEP-PKA (Chinese Event Prediction with Prompt-driven Knowledge Augmentation). Specifically, we design a prompt engineering strategy to leverage an LLM to generate fine-grained knowledge for event elements, including triggers and arguments. We then fine-tune a pre-trained language model (PLM) to learn a joint representation of events and its associated knowledge. Experimental results in two public Chinese benchmarks demonstrate that CEP-PKA outperforms baseline models relying solely on LLMs or PLMs, achieving average accuracy improvements of 10.62% and 12.51%, respectively.

Keywords: event prediction; knowledge augmentation; large language models



Academic Editor: Ming Liu

Received: 23 October 2025

Revised: 19 November 2025

Accepted: 24 November 2025

Published: 26 November 2025

Citation: Lou, S.; Xie, X.; Liu, W.; Jiang, W. Enhancing Chinese Event Prediction with Prompt-Driven Knowledge Augmentation. *Appl. Sci.* **2025**, *15*, 12543. <https://doi.org/10.3390/app152312543>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Events serve as fundamental units of human knowledge and are central to human comprehension of the world, making event prediction a critical task for natural language processing (NLP). Current event prediction methods largely depend on structured data, such as event pairs [1], event chains [2], or event graphs [3], all of which are fundamentally dependent on structured event datasets and require labor-intensive efforts to construct from unstructured text [4]. Moreover, most human-annotated datasets are limited in scale and consist of short, ambiguous, and sparse texts, bringing significant challenges for model training.

Knowledge augmentation has emerged as a promising approach for event prediction, enhancing both the interpretability of NLP models and the performance of downstream tasks. Existing methods [5,6] typically adopt a retrieval-and-integration paradigm, relying heavily on structured data, particularly knowledge graphs (KGs). This paradigm involves extracting event elements from text, retrieving relevant triples from KGs, and integrating them into event representations. While effective in structured knowledge scenarios, these approaches are inherently limited by the KG's scale and timeliness, etc. As depicted in Figure 1, taking the sentence “*Wawrinka has advanced to the next round of the 2019 US Open*

men's singles after beating Lorenzi 3–0 in the third round in New York.” as an example, event prediction based on KGs would face three limitations.

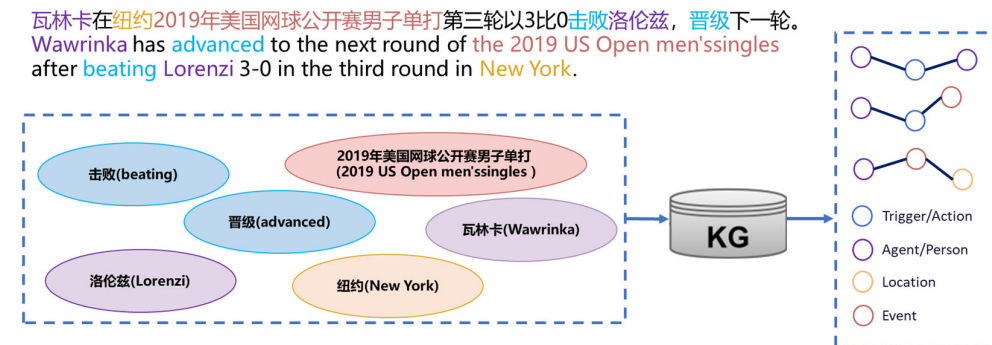


Figure 1. Traditional knowledge argumentation method for event prediction. This method generally has three steps: (1) extract event triggers and arguments from text-based events; (2) retrieve them in KGs; and (3) integrate retrieved triples with initial events.

- **Structural Rigidity:** KGs rely on pre-defined schemas that struggle to accommodate compound event elements. For instance, the phrase “2019 US Open men's singles” represents a specific, time-bound entity that may not fit into a KG's fixed entity schemas. Traditional KGs often lack flexible mechanisms to capture such context-specific or emerging entities, leading to an incomplete semantic representation of events.
- **Coverage Gaps:** Static KGs typically fail to include up-to-date information for the latest events. In this example, if the KG lacks entries for “2019 US Open men's singles”, it cannot retrieve the relevant background knowledge, resulting in an incomplete knowledge integration. This limitation is further evident in scenarios involving newly emerging events (e.g., “2023 US Open men's singles”) that have not been added to the KG.
- **Ambiguity Challenges:** KGs often struggle with entity disambiguation, particularly for ambiguous proper nouns. The name “Lorenzi” in the sentence could refer to multiple individuals, but a traditional KG may lack contextual clues to distinguish which specific “Lorenzi” is mentioned. Without additional disambiguation mechanisms, the KG might incorrectly associate the event with the wrong entity, introducing misleading knowledge into the prediction process and degrading accuracy.

These limitations of KG-based approaches highlight the need for more dynamic knowledge augmentation strategies. Recently, large language models have emerged as a revolutionary solution in knowledge engineering, with models like GPT-4 [7] demonstrating exceptional capabilities in natural language generation and context-aware reasoning. Unlike static KGs, LLMs can (1) generate fine-grained knowledge for novel entities (e.g., “2019 US Open men's singles”) by leveraging their vast pre-training data; (2) disambiguate ambiguous terms (e.g., “Lorenzi”) through contextual understanding, avoiding the pitfalls of KG-based entity confusion; and (3) dynamically adapt to evolving event schemas without relying on pre-defined structures.

However, LLMs are not without their own challenges, such as knowledge cutoffs and the propensity for hallucination, which can introduce factual inaccuracies into the knowledge augmentation process. The potential of large language models as proactive knowledge sources for event prediction has remained underexplored. To address this, we propose CEP-PKA (Chinese Event Prediction with Prompt-driven Knowledge Augmentation), a model that replaces traditional knowledge graphs by utilizing a prompt-driven framework for LLMs. (1) **Prompt Engineering:** We design custom templates to guide LLMs in generating context-specific knowledge for event triggers and arguments, which are fundamental

elements of event detection, thereby enriching their semantic depth. (2) **PLM Fine-Tuning:** A pre-trained language model is fine-tuned on LLM-augmented event texts, learning to integrate raw text with generated knowledge for an accurate event type classification. Experimental results demonstrate that CEP-PKA significantly outperforms LLM-only and PLM-only baselines, achieving up to 12.51% accuracy improvement in Chinese event benchmarks. By addressing the structural rigidity, coverage gaps, and ambiguity challenges of KGs, CEP-PKA establishes a new paradigm for knowledge-augmented event prediction, where LLMs serve as dynamic, scalable knowledge generators rather than passive retrieval systems.

Our contributions are summarized as follows:

- **Event Prediction with Fine-Grained Information:** We leverage explicit fine-grained information within event texts to enable accurate event predictions, addressing the sparsity challenge in short-text scenarios.
- **Prompt-Driven Augmentation:** We propose a prompt-driven framework using LLMs to generate context-specific knowledge for triggers/arguments, enhancing semantic richness and addressing short-text sparsity.
- **Systematic PLM Fine-Tuning:** We establish a robust training pipeline through rigorous fine-tuning of PLMs on two Chinese event prediction benchmarks.
- **Performance Validation:** Experiments show our method significantly improves prediction accuracy over standalone LLMs/PLMs, validating the efficacy of integrated knowledge augmentation.

2. Related Works

2.1. Knowledge Augmentation

In the field of event prediction, the limited scale of most datasets and the brevity of texts bring significant challenges. Previous studies [8,9] have demonstrated that supplementary knowledge can enhance semantic richness, mitigate data sparsity and ambiguity, and clarify contextual meaning, making knowledge augmentation a critical approach. Traditional methods employ retrieval-based strategies to acquire knowledge from knowledge graphs [10,11]. Although effective in structured scenarios, they are inherently constrained by rigid schemas and incomplete domain coverage.

Recent breakthroughs in large language models offer a paradigm shift for knowledge augmentation. Unlike static KGs, LLMs, which are trained on vast and diverse corpora, possess inherent capabilities for context-aware knowledge generation. While existing studies have focused on using external knowledge to refine prompt design [12,13], the potential of using prompts to extract knowledge from LLMs remains unexplored. In this work, we leverage LLMs' pre-trained knowledge and generative capabilities to dynamically predict supplementary event information, addressing traditional KGs' sparsity and ambiguity limitations via adaptive semantic enrichment. By treating LLMs as proactive knowledge sources, we aim to enhance event prediction performance through context-aware knowledge enrichment.

2.2. Event Prediction

Existing event prediction approaches predominantly rely on event graphs to model event arguments and cross-event relationships. Li et al. [14] constructed a narrative event evolution graph using scaled graph neural networks (GNNs) for event prediction. Liu et al. [15] proposed a directed labeled syntactic information graph to enhance GNN capabilities for event factuality prediction. Yu et al. [16] developed an online neural symbolic framework for temporal knowledge graph forecasting. Mathur et al. [17] created a weakly labeled event logic graph for document-level script event prediction.

While these graph-based methods have demonstrated effectiveness, their applicability is largely limited to structured event graphs. However, most real-world event data exists in unstructured text, and converting all text-based events into graph structures is both labor-intensive and impractical. To address this gap, our work focuses on direct event type prediction from raw text, eliminating the need for manual graph construction.

3. Problem Definition

Event prediction involves analyzing the information presented in a given text, identifying the underlying event, and determining its type from a predefined set of event types. Given the nature of this task, we model event prediction as a textual event classification problem. Formally, given N textual inputs $X = \{x_1, x_2, \dots, x_N\}$, the task aims to assign each input x a label y from a fixed label set $Y = \{y_1, y_2, \dots, y_M\}$, which contains M distinct event types. Each label y represents a specific type of event, and the goal is to accurately map the input text to the most appropriate event type, thereby predicting the event that the text describes.

4. Methodology

The CEP-PKA framework comprises three core modules: Fine-grained event decomposition, prompt construction, and fine-tuning (Figure 2). In the fine-grained event decomposition module, events are decomposed into granular lexical units, which serve as precise targets for knowledge augmentation to tackle semantic sparsity in short-text scenarios. In the prompt construction module, a tailored prompt framework is designed to generate context-aware knowledge for both event triggers and arguments, effectively bridging semantic gaps and enhancing the representation of event components. In the fine-tuning module, a pre-trained language model is rigorously fine-tuned on two Chinese event prediction datasets to optimize model performance for event type classification.

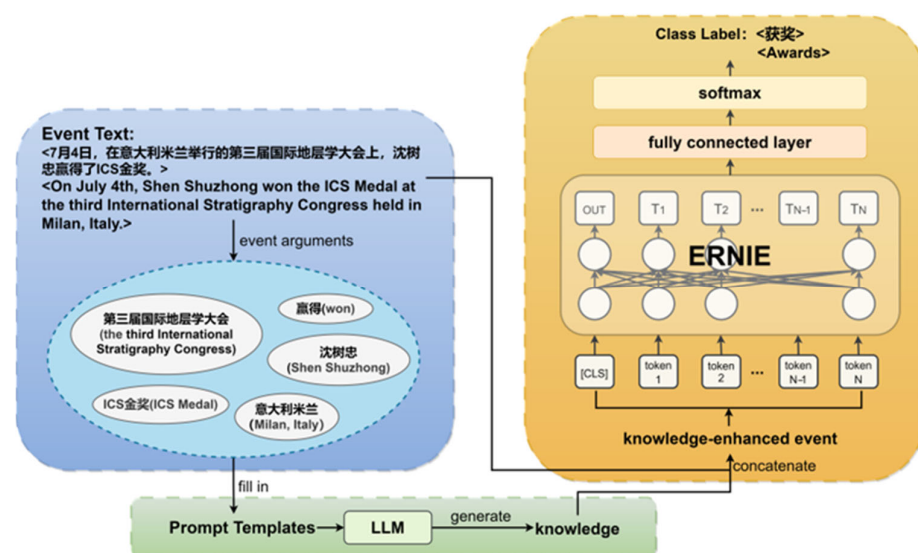


Figure 2. The architecture of CEP-PKA.

4.1. Fine-Grained Event Decomposition

To enhance the model's comprehension of the explicit information within event texts, we leverage fine-grained details from the original texts. This strategy is derived from the observation that event-related texts usually contain some specific linguistic elements, which are crucial for capturing the semantics of the described events. Specifically, the fine-

grained event information emphasizes event trigger and event arguments, as these elements represent the most prominent and informative components in the event description.

Event triggers, which mostly appear in the form of verbs or verb phrases, accurately capture the core actions related to the event and determine the basic attributes and core behaviors of the event. For example, in the sentence “On 4 July, Shen Shuzhong won the ICS Medal at the third International Stratigraphy Congress held in Milan, Italy.”, the word “won” is the core trigger word of the event, which clearly expresses the main action of the event.

Event arguments provide rich, detailed information for the event, covering key dimensions such as entities, time, and location in the event. These arguments constitute important background information for understanding the event. Taking the above sentence as an example, “Shen Shuzhong” is the main entity of the event, “ICS Award” is the entity receiving the award, “the Third International Congress of Stratigraphy” is the background venue of the event, and “Milan, Italy” is the geographical location of the event. These arguments are interrelated, providing a multi-dimensional basis for judging the event type and helping the model comprehensively capture the information of the event.

Formally, for each event e_i , A_i is the set of its event triggers and event arguments:

$$A_i = \{a_1, a_2, \dots, a_j, \dots, a_n\} \quad (1)$$

where n is the total number of event arguments for event e_i , and each a_j ($1 \leq j \leq n$) denotes an individual argument (e.g., trigger, agent, time, or location).

4.2. Prompt Construction

To address the critical need for semantic expansion of original event representations and integration of background knowledge and commonsense information, we have developed a sophisticated set of prompt engineering templates. These templates are strategically designed to direct large language models to generate detailed, contextually relevant knowledge that closely aligns with fine-grained event elements. As a result, they bridge the semantic gaps and enhance the comprehensibility of event semantics.

For the acquisition of external knowledge related to event elements, we employ an open-ended question template. This approach effectively guides LLMs to generate rich, pertinent information for fine-grained event elements. Leveraging GPT-3.5, one of the most advanced LLMs, we are able to obtain detailed and high-quality descriptions. The specific templates used in our research are as follows:

- For Event Triggers: Please provide a detailed explanation of the meaning and function of the event trigger [*trigger word*] in the context.
- For Entity Arguments: Please provide background information about the entity [*entity word*], including its type, main characteristics, and related events.
- For Time Arguments: Please describe typical events or background related to the time expression [*time words*].
- For Location Arguments: Please introduce the nature and importance of the location [*location words*], as well as the types of events that commonly occur there.

For example, consider the sentence above, by applying the templates presented above, an LLM can generate rich background knowledge for each event element, as illustrated in Figure 3.



Figure 3. The examples of prompts and answers for knowledge generation.

Furthermore, we design a single-choice question-based prompt template for direct event prediction by presenting an LLM with a target event and requiring it to select the most appropriate event type from a predefined label set. Essentially, the predictions made by the LLM follow a zero-shot learning paradigm (e.g., “The event type of [text] is: [Option 1, Option 2, ...].”), as presented in Figure 4.

Formally, we fill in the prompt template with all the elements in A_i to generate relevant knowledge K_i through the LLM:

$$K_j = \text{LLM}(\text{Prompt}(a_j)) \quad (2)$$

$$K_i = \{k_1, k_2, \dots, k_j, \dots, k_n\} \quad (3)$$

where K_j represents the knowledge associated with a_j , expressed as a sentence or paragraph recording specific contextual information, and K_i denotes the aggregate knowledge of e_i .

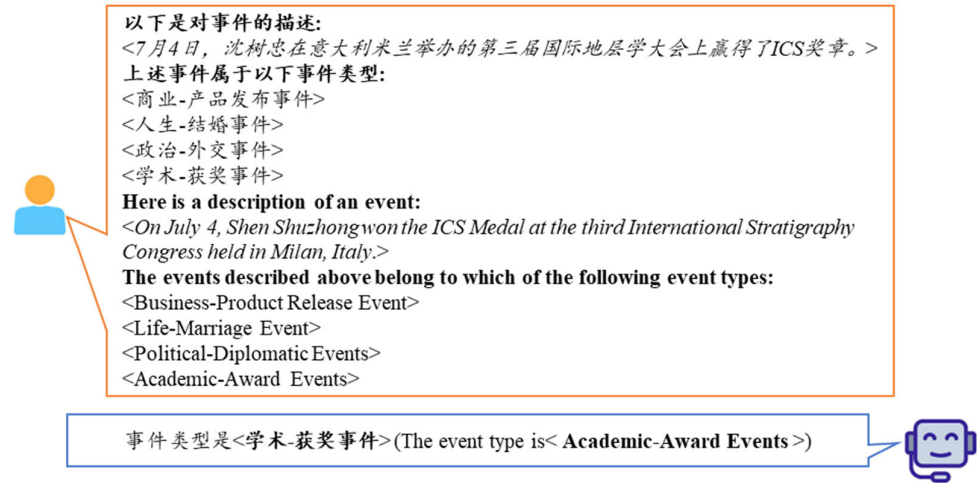


Figure 4. The examples of prompts and answers for event prediction.

4.3. Fine-Tuning

We employ ERNIE (Enhanced Representation through Knowledge Integration) [18] as our backbone pre-trained language model for the final event classification task.

ERNIE is a series of large-scale pre-trained models developed by Baidu, specifically designed for Chinese natural language processing. It enhances semantic representation by integrating knowledge from various sources during pre-training, making it particularly effective for understanding Chinese text. We choose ERNIE due to its proven superiority in Chinese NLP benchmarks.

We fine-tune ERNIE using the knowledge-augmented event texts generated in the preceding step. Specifically, the input text is formed by concatenating e_i and its corresponding k_i using special tokens to clearly demarcate different information sources. The final input format is as follows:

$$x = [\text{CLS}] + e_i + [\text{SEP}] + k_1 + [\text{SEP}] + k_2 + \dots + k_i + [\text{SEP}] \quad (4)$$

This structured format helps the ERNIE model distinguish the original event text from the LLM-generated knowledge, learning to integrate them effectively. Subsequently, the input is tokenized and fed into the ERNIE model to generate contextualized embeddings, which are then processed by a pooling layer. The output of the pooling layer is transmitted through a fully connected layer, which maps the high-dimensional embedding to the specified number of output classes:

$$y = Wx + b \quad (5)$$

where W denotes an $W \times H$ matrix, M represents the number of output classes, and H denotes the hidden layer size. Additionally, b is a bias vector of P -dimension, and y is the output vector, also of P -dimension.

To predict the final event type, the logits are passed through a softmax function as the activation function to obtain the probabilities of each class:

$$\sigma(z)_p = \frac{e^{z_p}}{\sum_{m=1}^M e^{z_m}} \quad (6)$$

where $\sigma(z)_p$ is the p -th ($1 \leq p \leq M$) component of the output vector after applying the softmax function, z_p is the p -th component of the input vector z , and M is the number of classes.

Given that our task is essentially a multi-class classification problem, we employ cross-entropy loss, a loss function well-adapted for scenarios where each instance belongs to exactly one of the multiple classes:

$$L = \sum_m y_m \log \hat{y}_m \quad (7)$$

where y_m is the true label for class m and $\hat{y}_m$ is the predicted probability of class m .

Similarly to Bert, we refrain from optimizing bias and parameters and apply the BERTAdam optimizer to the remaining parameters. A warm-up schedule is adopted, with the learning rate defined as follows:

$$l_{r_t} = \begin{cases} \frac{t}{N_{warmup}} & \text{for } t < N_{warmup} \\ \max\left(\frac{t-1}{N_{warmup}-1}, 0\right) & \text{for } t \geq N_{warmup} \end{cases} \quad (8)$$

where l_{r_t} is the learning rate at time t and N_{warmup} is the total number of warm-up steps.

The parameter update with weight decay is defined as:

$$\theta_t = \theta_{t-1} - l_{r_t} \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda \theta_{t-1} \right) \quad (9)$$

where l_{r_t} is the learning rate at time step t , $\hat{m}_t$ represents the mean of gradients after bias correction, $\hat{v}_t$ represents the variance of gradients after bias correction, and λ is a fixed-weight decay coefficient.

5. Experiment

5.1. Implementation Details

Datasets: We primarily evaluate our framework on two established Chinese event prediction benchmarks: DuEE [19] and Title2Event [20], and further conduct pilot experiments on the more recent LEVEN dataset [21] to assess generalizability in specialized domains.

DuEE (Baidu's DuEE Dataset) is a large-scale event extraction corpus containing 19,640 events with 65 types across domains such as finance, sports, and public safety. Each instance includes a sentence with annotated trigger and argument spans, making it suitable for evaluating both detection and classification. The dataset was constructed from news articles published before 2021.

Title2Event focuses exclusively on short-text scenarios using news headlines (average length: 23 characters) as input. Following Deng's approach [20], we group the 10 most frequent classes in Title2Event and label the rest as "Other Topics". Notably, since the events in the datasets occurred before 2022 and GPT-3.5 (trained on data up to January 2022) is used to generate temporal knowledge, we posit that it can effectively provide accurate and context-relevant information.

To assess generalizability to specialized domains and more recent data, we also conduct pilot experiments on LEVEN [21], a large-scale Chinese legal event detection dataset released in 2022. LEVEN comprises over 100,000 legal documents with 108 fine-grained event types (e.g., "Contract Dispute", "Administrative Penalty") and rich argument structures. Notably, all datasets used in our study—including LEVEN—predate GPT-3.5's knowledge cutoff (early 2022), ensuring no temporal data leakage during knowledge generation. Although its annotation schema differs slightly from DuEE, we adapt our framework by mapping triggers to LEVEN's event definitions. On a randomly sampled test subset ($n = 1000$), CEP-PKA improves F1 by +2.3% over vanilla ERNIE, demonstrating robustness in domain-specific settings. Full integration is reserved for future work.

As shown in Table 1, the length distribution of event texts changes significantly after knowledge enhancement. This indicates that knowledge integration effectively enriches event content while mitigating the occurrence of excessively long texts.

Table 1. The text length distribution of the initial events and the augmented events with knowledge in the DuEE and Title2Event.

	DuEE w/o Knowledge	DuEE w/Knowledge	Title2Event w/o Knowledge	Title2Event w/Knowledge
Max/Min/Average Length	2554/8/58	1012/12/196	50/7/226	780/8/244
1st/2nd/3rd Quartile	26/41/79	108/177/269	22/27/30	144/230/324

Baselines: We compare CEP-PKA against widely used text classification models, including (1) BERT-family models: BERT, BERT + CNN/RNN/RCNN/DPCNN, RoBERTa [22] and MacBERT [23]; (2) enhanced models: Label Confusion Model(LCM) [24], Distribution Calibration(DC) [25], SAtt-LSTM [26], CED-EOSN [27], EDLA [28], miniRBT [29], PMTNet [30]; (3) knowledge graph (KG)-enhanced models: KG-BERT + ERNIE [31] and KEPLER-Chinese [32].

Configuration: The padding size is set to 256 for original text-based events and adjusted to 512 for knowledge-augmented events. All experiments employ a learning rate of 0.00005 and a hidden layer size of 768, with each experiment run 10 times and the average taken as the final result; the LLM version used is gpt-3.5-turbo.

5.2. Main Results

Comparative results (Table 2) demonstrate that CEP-PKA outperforms all baseline models across all evaluation metrics on both DuEE and Title2Event datasets, validating its effectiveness. Specifically, CEP-PKA achieves an event prediction accuracy of 0.9463 on DuEE, surpassing the best baseline by 6.76%, and 0.8820 on Title2Event, outperforming baselines with an accuracy range from 0.72 to 0.79. This superiority stems from two key factors: (1) integration of ERNIE, a PLM optimized for the Chinese corpus, which enhances language understanding; and (2) incorporation of LLM-generated knowledge, which supplements contextual information and enriches semantic representations. (3) CEP-PKA maintains distinct advantages over knowledge-enhanced baselines (K-BERT, KEPLER) in event prediction. On DuEE, it outperforms K-BERT (Macro-F1 0.8420, accuracy 0.8651) and KEPLER (Macro-F1 0.9302, accuracy 0.9407) with its Macro-F1 of 0.9355 and accuracy of 0.9463. On Title2Event (short, context-sparse texts), the gap widens—CEP-PKA's 0.8820 accuracy far exceeds K-BERT's 0.7809 and KEPLER's 0.8204. The key reason why CEP-PKA outperforms K-BERT and KEPLER lies in the fact that, unlike K-BERT and KEPLER which use fixed, pre-built knowledge graphs, CEP-PKA leverages dynamic knowledge generated by large language models. This dynamic knowledge better captures the context-specific semantics required for event prediction across diverse datasets.

Table 2. Comparison of CEP-PKA baseline models. † signifies the results are taken from their original papers and * means that the results are achieved with their released code under default settings.

Dataset	Method	Macro			Weighted			ACC
		P	R	F1	P	R	F1	
DuEE	CEP-PKA	0.9387	0.9374	0.9355	0.9492	0.9463	0.9461	0.9463
	Bert *	0.8164	0.8304	0.8138	0.8815	0.8754	0.8746	0.8754
	Bert + CNN *	0.8138	0.8080	0.7987	0.8700	0.8619	0.8608	0.8619
	Bert + RNN *	0.7320	0.7402	0.7226	0.8335	0.8478	0.8351	0.8478
	Bert + RCNN *	0.8461	0.8523	0.8415	0.8844	0.8787	0.8779	0.8787
	Bert + DPCNN*	0.6754	0.6207	0.6099	0.7950	0.7843	0.7755	0.7843
	Roberta *	0.7849	0.6744	0.7001	0.8059	0.8022	0.7905	0.8022
	Macbert *	0.8080	0.7261	0.7480	0.8347	0.8305	0.8237	0.8305
	LCM *	0.7863	0.7736	0.7799	0.8400	0.8503	0.8451	0.7879
	DC *	0.4785	0.5218	0.4726	0.6483	0.5307	0.5567	0.5307
	SAtt-LSTM †	0.6910	0.8120	0.7310	-	-	-	-
	CED-EOSN †	0.9350	0.9110	0.9230	-	-	-	-
	EDLA †	0.9439	0.9726	0.9581	-	-	-	-
	miniRBT *	0.4986	0.32094	0.29196	0.74826	0.8667	0.68328	0.8667
	PMTNet †	0.7629	0.7797	0.7712	-	-	-	-
	K-BERT *	0.8352	0.8505	0.8420	0.8608	0.8755	0.8673	0.8651
	KEPLER *	0.9205	0.9407	0.9302	0.9356	0.9503	0.9421	0.9407
Title2Event	CEP-PKA	0.8787	0.8866	0.8811	0.8863	0.8820	0.8829	0.8820
	Bert *	0.7989	0.7116	0.7434	0.7730	0.7640	0.7619	0.7640
	Bert + CNN *	0.7927	0.7625	0.7746	0.7890	0.7859	0.7839	0.7859
	Bert + RNN *	0.7264	0.6830	0.6976	0.7269	0.7241	0.7213	0.7241
	Bert + RCNN *	0.7691	0.7799	0.7709	0.7793	0.7775	0.7755	0.7775
	Bert + DPCNN*	0.6588	0.6622	0.6543	0.7210	0.7295	0.7226	0.7295
	Roberta *	0.7785	0.7316	0.7513	0.7539	0.7534	0.7487	0.7534
	Macbert *	0.7882	0.7430	0.7618	0.7642	0.7637	0.7593	0.7637
	DC *	0.4602	0.6009	0.4792	0.5779	0.4652	0.4552	0.4652
	LCM *	0.7698	0.7909	0.7802	0.7845	0.8222	0.8029	0.7534
	miniRBT *	0.7332	0.6571	0.6872	0.6972	0.6975	0.6894	0.6975
	K-BERT *	0.7854	0.7608	0.7723	0.7905	0.7808	0.7854	0.7809
	KEPLER *	0.8207	0.8003	0.8106	0.8305	0.8201	0.8250	0.8204

5.3. Ablation Study and Further Analysis

The impact of LLM-generated knowledge: As shown in Table 3, we evaluate the contribution of introduced knowledge to CEP-PKA on DuEE and Title2Event. Results indicate that most evaluation metrics across both datasets show significant improvements with knowledge integration, underscoring the positive impact of LLM-generated knowledge on CEP-PKA's performance. Despite potential LLM hallucinations, findings confirm that structured knowledge from LLMs meaningfully enhances downstream task accuracy. We hypothesize that this is because the fine-tuned ERNIE model learns robust semantic representations that can align with and filter out highly relevant knowledge, making it resilient to occasional noise or inconsistencies from LLM outputs.

Comparison of Different LLMs for Knowledge Generation: To assess the generalization ability of CEP-PKA across different language models, we replaced the GPT-3.5-Turbo used in knowledge generation with two representative Chinese LLMs, Qwen-Turbo and DeepSeek-V3.2-Exp, while keeping all other experimental settings identical. As summarized in Table 4, CEP-PKA achieves consistently strong performance on both DuEE and T2E datasets under all LLMs. On DuEE, GPT-3.5-Turbo yields the highest overall accuracy (0.9463), whereas Qwen and DeepSeek remain highly competitive, with F1 scores within 1% of the baseline. On T2E, DeepSeek-V3.2-Exp slightly outperforms GPT-3.5. These findings

demonstrate that CEP-PKA’s improvement primarily stems from its prompt-driven knowledge augmentation mechanism rather than dependence on a specific LLM, confirming the framework’s model-agnostic adaptability. Moreover, Qwen and DeepSeek achieve comparable accuracy with substantially lower inference cost, indicating their potential as cost-effective alternatives to commercial APIs. This observation also motivates further exploration of lightweight and distilled models, as discussed in Section 5.4, to enhance the scalability and practicality of LLM-augmented event prediction.

Table 3. The influence of knowledge augmentation on CEP-PKA. ✓ denotes the presence of a specific component and × denotes the absence.

Dataset	Knowledge	Macro			Weighted			ACC
		P	R	F1	P	R	F1	
DuEE	×	0.9021	0.8971	0.8906	0.9370	0.9351	0.9329	0.9351
	✓	0.9387	0.9374	0.9455	0.9492	0.9463	0.9461	0.9463
		(+3.66%)	(+4.03%)	(+4.49%)	(+1.22%)	(+1.12)	(+1.32%)	(+1.12%)
Title2Event	×	0.8812	0.8837	0.8799	0.8804	0.8769	0.8765	0.8769
	✓	0.8787	0.8866	0.8811	0.8863	0.8820	0.8829	0.8820
		(−0.25%)	(+0.29%)	(+0.12%)	(+0.59%)	(+0.51%)	(+0.64%)	(+0.51%)

Table 4. Performance comparison of CEP-PKA using different large language models (LLMs).

Dataset	LLMs	Macro			Weighted			ACC
		P	R	F1	P	R	F1	
DuEE	Qwen	0.9315	0.9220	0.9242	0.9409	0.9388	0.9382	0.9388
	Deepseek	0.9228	0.9342	0.9249	0.9412	0.9366	0.9371	0.9366
	GPT 3.5	0.9387	0.9374	0.9355	0.9492	0.9463	0.9461	0.9463
Title2Event	Qwen	0.8817	0.8492	0.8629	0.8729	0.8715	0.8708	0.8715
	Deepseek	0.8780	0.8881	0.8814	0.8859	0.8827	0.8830	0.8827
	GPT 3.5	0.8787	0.8866	0.8811	0.8863	0.8820	0.8829	0.8820

The impact of different PLMs: Figure 5 compares various PLMs using CEP-PKA’s fine-tuning pipeline. CEP-PKA consistently outperforms other models across all metrics on both datasets. This advantage is attributed to ERNIE’s training on a Chinese corpus, which enables stronger semantic representation and event prediction capabilities after fine-tuning, highlighting the critical role of language-specific pre-training in cross-lingual NLP tasks.

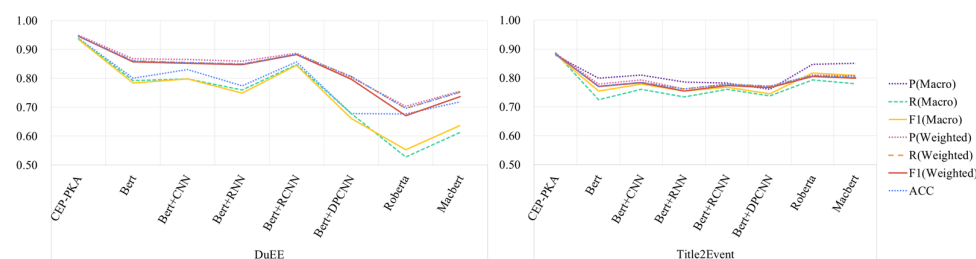


Figure 5. The influence of different PLMs on CEP-PKA.

Comparison of knowledge retrieval capabilities between LLM and KGs: We assess the knowledge retrieval performance of two large-scale Chinese KGs by calculating the success rate of retrieving event argument knowledge, compared with knowledge generated by GPT-3.5, for all events in DuEE. As illustrated in Figure 6, GPT-3.5 demonstrates significantly stronger capabilities in generating knowledge for event arguments: it outperforms

KGs in coverage, as the latter can only retrieve knowledge for a limited subset of arguments, primarily constrained by their dataset scale. Moreover, GPT-3.5 demonstrates superior ability to interpret compound nouns within event arguments and generate relevant knowledge, whereas KG-based methods often face challenges with such complex linguistic structures.



Figure 6. Percentage of matched knowledge in different knowledge sources.

Comparison of CEP-PKA with prompt-driven GPT: We conduct a comprehensive performance comparison between the prompt-driven GPT-3.5 under a zero-shot learning paradigm and our proposed CEP-PKA method. As shown in Table 5, CEP-PKA outperforms the GPT-3.5-based prompt approach on both datasets, even when GPT-3.5 incorporates its self-generated knowledge, highlighting the effectiveness of CEP-PKA in Chinese event prediction tasks.

Table 5. Comparison of CEP-PKA and direct LLM-based prediction. ✓ denotes the presence of a specific component and × denotes the absence.

Dataset	Model	Knowledge	Macro			Weighted			ACC
			P	R	F1	P	R	F1	
DuEE	CEP-PKA	✓	0.9387	0.9374	0.9455	0.9492	0.9463	0.9461	0.9463
	GPT3.5	✓	0.7829	0.8249	0.7732	0.8478	0.7590	0.7709	0.7590
		×	0.7711	0.8042	0.7549	0.8455	0.7522	0.7680	0.7522
Title2Event	CEP-PKA	✓	0.8787	0.8866	0.8811	0.8863	0.8820	0.8829	0.8820
	GPT3.5	✓	0.5824	0.6888	0.5752	0.6326	0.5266	0.4702	0.5266
		×	0.5884	0.6668	0.5625	0.6183	0.5184	0.4721	0.5184

To further evaluate CEP-PKA's stability and robustness, we tested its performance across 65 event types in DuEE. As illustrated in Figure 7, CEP-PKA consistently achieves high scores (marked by red blocks), while GPT-3.5's performance (marked by cool-toned blocks) is less consistent and lower on average. These results confirm that CEP-PKA delivers stronger and more balanced performance across all event types, demonstrating its capability to handle diverse scenarios effectively.

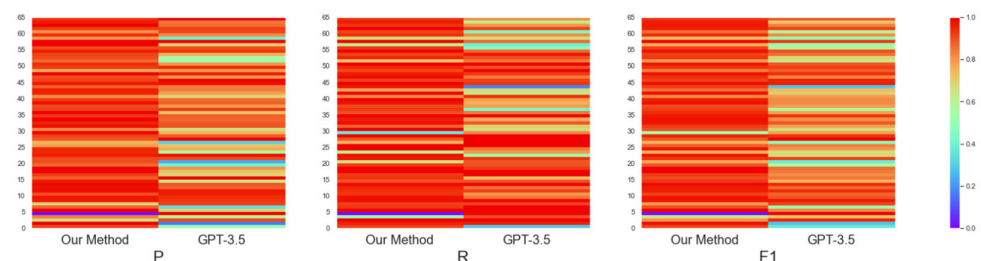


Figure 7. Heat maps of the prediction performance for different event types in DuEE. The left columns in all maps are from CEP-PKA, while the right ones are from GPT-3.5. The vertical axis is the 65 event types in DuEE.

The impact of hyperparameters: We study the setting of parameters (learning rate, random seed, padding size) and their effects on the model.

On the DuEE dataset, we first systematically investigate the impact of learning rates on the performance of CEP-PKA. Through extensive hyperparameter tuning, we determine

that a learning rate of 0.00005 yields optimal results, outperforming other tested rates in terms of both prediction accuracy and loss minimization, as vividly illustrated in Figure 8a.

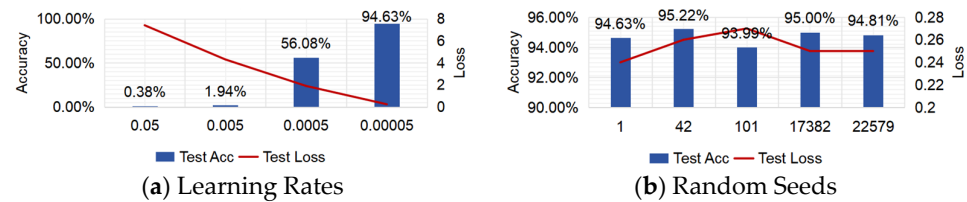


Figure 8. Impact of learning rates and random seeds on the performance of CEP-PKA on DuEE.

Experiments on DuEE with varying random seeds (Figure 8b) demonstrate minimal variance in accuracy, confirming CEP-PKA’s robustness to different weight initializations. This consistency underscores the model’s reliability in real-world deployments.

We also explored the influence of padding size on model performance. First, we analyzed the length distribution of original event texts and compared it with that after knowledge augmentation. As shown in Figure 9, for the DuEE and Title2Event datasets, only 1.7% and 5.0% of the data, respectively, exceed the 512-character padding limit. Considering that the maximum input length supported by the ERNIE model is 512, we strategically set the padding size to 512 to cover the vast majority of text instances.

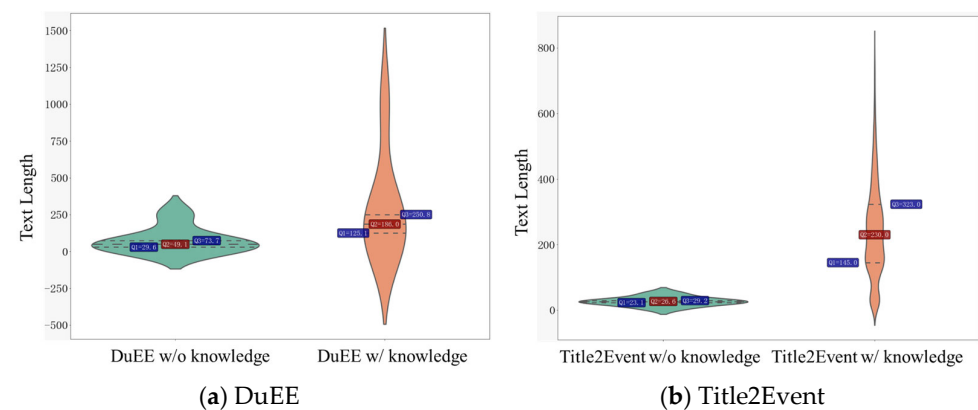


Figure 9. The length distribution of the initial event text and the event text after knowledge augmentation. The three dotted lines in the figure represent the first quartile (Q1, 25%), median (Q2, 50%), and the third quartile (Q3, 75%).

Finally, with the random seed fixed at 42, we comprehensively evaluated the impact of different padding sizes on model performance. The experimental results in Figure 10 show that the CEP-PKA model achieves the highest prediction performance when the padding size is set to 512. As the padding size decreases, the model’s performance gradually approaches that of relying solely on the original event text.

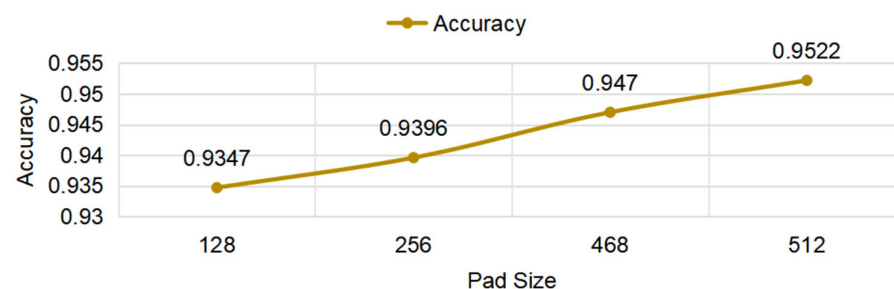


Figure 10. Impact of pad sizes on the performance of CEP-PKA on DuEE.

5.4. Limitation

Despite the promising performance of our CEP-PKA framework, its reliance on Large Language Models (LLMs)—particularly commercial APIs like GPT-3.5—for knowledge augmentation introduces several inherent limitations that affect both reliability and practicality.

First, LLMs are susceptible to hallucinations, i.e., generating factually inaccurate or fabricated content. When such misleading knowledge is injected into the event prediction pipeline, it risks distorting semantic representations or steering the classifier toward incorrect predictions. Second, LLMs suffer from temporal knowledge cutoffs. For instance, GPT-3.5's training data ends in June 2023, rendering it unable to accurately describe events, organizations, or policies that emerged afterward (e.g., China's 2024 AI regulatory framework). While this limitation does not bias our current evaluation—since all datasets (DuEE, Title2Event, and LEVEN) were constructed prior to 2023—it poses a significant challenge for real-world deployment in dynamic environments.

Moreover, the heavy dependence on closed-source, commercial LLM APIs incurs substantial computational and financial costs, especially when scaling to large datasets. This hinders reproducibility, limits accessibility for resource-constrained researchers, and impedes practical adoption in production systems.

To address these challenges, we propose several actionable strategies for future work:

- **Knowledge Verification:** Integrate an automated validation layer that cross-checks LLM-generated knowledge against trusted external sources (e.g., CN-DBpedia, Own-Think) before augmentation.
- **Confidence-based Filtering:** Leverage the LLM's internal confidence scores to discard low-certainty or ambiguous generations.
- **Ensemble Knowledge Generation:** Aggregate outputs from multiple diverse LLMs (e.g., GPT-3.5, Qwen, GLM, Baichuan) to identify consensus facts and suppress outliers.
- **Knowledge Distillation:** Train a lightweight student model (e.g., MiniRBT) to mimic the behavior of the full CEP-PKA pipeline, preserving performance while eliminating recurring API costs.
- **Domain-specific Fine-tuning:** Fine-tune LLMs on fact-checked, domain-relevant corpora to enhance factual consistency and reduce hallucination rates.

Notably, our results in Table 4 already provide empirical support for a cost-effective alternative. Open-source models such as Qwen and DeepSeek achieve performance comparable to GPT-3.5 at a fraction of the cost. This validates that strategies like ensemble generation and distillation are not merely theoretical—they offer a practical, scalable pathway for deploying CEP-PKA in real-world Chinese event prediction systems.

Ultimately, improving both the factual reliability and computational efficiency of LLM-augmented frameworks is essential. For dynamic applications, combining CEP-PKA with retrieval-augmented generation (RAG) or using continually updated open-source LLMs (e.g., Qwen-Max, DeepSeek-V3) represents a promising direction to mitigate knowledge decay and ensure long-term viability.

5.5. Case Study

For a qualitative analysis, we analyze two cases in DuEE to evaluate the semantic comprehension capabilities of GPT and CEP-PKA, as shown in Figure 11. Our observations reveal that GPT encounters difficulties in semantic comprehension. In Case 1, GPT misclassified the event as “Life-Childbirth” due to overemphasizing the phrase “becoming a mother” while ignoring the core action “visiting”; CEP-PKA correctly identified the event type by integrating global context. In Case 2, GPT focused narrowly on the word “hit” and incorrectly predicted “Disaster/Accident-Attack”, failing to recognize the event's ultimate outcome of “death”. These examples demonstrate CEP-PKA's ability to capture holistic

semantics rather than relying on isolated keywords, highlighting its robustness against semantic interference.

Case 1	Case 2
[Event] 10月14日，有媒体拍到高圆圆带着女儿探班赵又廷，而这也是高圆圆升级当妈后首次被拍到带娃外出。 <i>Some media captured Gao Yuanyuan visiting Zhao Youting with her daughter, and this is also the first time Gao Yuanyuan has been photographed taking her child out after becoming a mother.</i>	[Event] 7月2日下午，一名39岁的中年妇女遭到10岁男孩抛下的灭火器砸中，当场就不治身亡，而更加让人感到害怕的是，事发之前还有一个灭火器从高空抛下。 <i>On the afternoon of July 2, a 39-year-old middle-aged woman was hit by a fire extinguisher thrown by a 10-year-old boy and died on the spot. What's even more frightening is that another fire extinguisher was thrown from a high altitude before the incident.</i>
[Ground Truth] 交往-探班 Dating-Visiting [GPT 3.5] 人生-产子/女 Life-Childbirth ✖ [Our model] 交往-探班 Dating-Visiting ✔	[Ground Truth] 人生-死亡 Life-Death [GPT 3.5] 灾害/意外-袭击 Disaster Accident-Attack ✖ [Our model] 人生-死亡 Life-Death ✔

Figure 11. Case study of the outputs from CEP-PKA and GPT3.5.

5.6. Error Analysis

To gain deeper insights into our model's behavior, we analyzed cases where CEP-PKA made incorrect predictions. In one instance, the event “The enterprise announces corporate layoffs.” was misclassified as “The enterprise announces a corporate merger and acquisition”. Our analysis revealed that the LLM-generated knowledge for the entity “Corporate” was overly generic, focusing on its business nature rather than the specific context of a workforce reduction. This led to semantic confusion during fine-tuning. This case highlights the importance of generating context-specific knowledge and suggests a future improvement: incorporating a relevance-checking mechanism for the generated knowledge before integration.

6. Conclusions

In this work, we introduce CEP-PKA to enhance the accuracy of Chinese event prediction via prompt-driven knowledge augmentation. The approach employs large language models as knowledge sources to generate fine-grained information for event elements, which is then integrated into a pre-trained language model for event type prediction via fine-tuning. Experimental results demonstrate that CEP-PKA significantly outperforms baseline models, highlighting the potential of LLMs to enrich event text semantics. For future research, we plan to integrate diverse LLMs to enhance the diversity and richness of generated knowledge and extend the framework to analyze event causality and temporal event prediction, expanding its application to more complex event understanding scenarios.

Author Contributions: Conceptualization, W.L.; Data curation, W.J.; Methodology, S.L. and W.L.; Validation, S.L., X.X. and W.J.; Writing—original draft, S.L. and W.L.; Writing—review and editing, X.X. and W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China Major Projects: grant number 61991410.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to the need for subsequent scientific work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Chambers, N.; Jurafsky, D. Unsupervised learning of narrative event chains. In *Annual Meeting of the Association for Computational Linguistics, Proceedings of the ACL-08: HLT, Columbus, OH, USA, 15–20 June 2008*; Stanford University: Stanford, CA, USA, 2008; pp. 789–797. Available online: <https://api.semanticscholar.org/CorpusID:529375> (accessed on 23 November 2025).
2. Lv, S.; Qian, W.; Huang, L.; Han, J.; Hu, S. Sam-net: Integrating event-level and chain-level attentions to predict what happens next. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (2019)*, Honolulu, HI, USA, 27 January–1 February 2019. Available online: <https://api.semanticscholar.org/CorpusID:57983897> (accessed on 23 November 2025).
3. Huai, Z.; Zhang, D.; Yang, G.; Tao, J. Spatial-temporal knowledge graph network for event prediction. *Neurocomputing* **2023**, *553*, 126557. [\[CrossRef\]](#)
4. Zhao, L. Event Prediction in Big Data Era: A Systematic Survey. *ACM Comput. Surv. (CSUR)* **2020**, *54*, 1–37. [\[CrossRef\]](#)
5. Cao, Y.; Tang, Y.; Du, H.; Xu, F.; Wei, Z.; Jin, C. Heterogeneous reinforcement learning network for aspect-based sentiment classification with external knowledge. *IEEE Trans. Affect. Comput.* **2023**, *14*, 3362–3375. [\[CrossRef\]](#)
6. Gardères, F.; Ziaeeefard, M.; Abeloos, B.; Lécué, F. Conceptbert: Concept-aware representation for visual question answering. In *Findings of the Association for Computational Linguistics: EMNLP*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 489–498. Available online: <https://api.semanticscholar.org/CorpusID:226284018> (accessed on 23 November 2025).
7. Achiam, O.J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. GPT-4 Technical Report. 2023. Available online: <https://api.semanticscholar.org/CorpusID:257532815> (accessed on 23 November 2025).
8. Melamud, O.; Goldberger, J.; Dagan, I. context2vec: Learning generic context embedding with bidirectional LSTM. In *Proceedings of the 20th Conference on Computational Natural Language Learning (CoNLL 2016)*, Berlin, Germany, 11–12 August 2016. Available online: <https://api.semanticscholar.org/CorpusID:7890036> (accessed on 23 November 2025).
9. Huang, L.; Sun, C.; Qiu, X.; Huang, X. Glossbert: Bert for word sense disambiguation with gloss knowledge. *arXiv* **2019**, arXiv:1908.07245. Available online: <https://api.semanticscholar.org/CorpusID:201103745> (accessed on 23 November 2025).
10. Natu, S.; Sural, S.; Sarkar, S. External commonsense knowledge as a modality for social intelligence question-answering. In *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Paris, France, 2–6 October 2023; pp. 3036–3042. Available online: <https://api.semanticscholar.org/CorpusID:266151538> (accessed on 23 November 2025).
11. Bayat, F.F.; Qian, K.; Han, B.; Sang, Y.; Belyi, A.; Khorshidi, S.; Wu, F.; Ilyas, I.; Li, Y. FLEEK: Factual error detection and correction with evidence retrieved from external knowledge. *arXiv* **2023**, arXiv:2310.17119v1. Available online: <https://api.semanticscholar.org/CorpusID:264490676> (accessed on 23 November 2025). [\[CrossRef\]](#)
12. Liu, J.; Yang, L. Knowledge-enhanced prompt learning for few-shot text classification. *Big Data Cogn. Comput.* **2024**, *8*, 43. [\[CrossRef\]](#)
13. Yan, Y.; Zheng, P.; Wang, Y. Enhancing large language model capabilities for rumor detection with knowledge-powered prompting. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108259. [\[CrossRef\]](#)
14. Li, Z.; Ding, X.; Liu, T. Constructing narrative event evolutionary graph for script event prediction. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, Stockholm, Sweden, 13–19 July 2018. Available online: <https://api.semanticscholar.org/CorpusID:21723549> (accessed on 23 November 2025).
15. Liu, X.; Huang, H.; Zhang, Y. End-to-end event factuality prediction using directional labeled graph recurrent network. *Inf. Process. Manag.* **2022**, *59*, 102836. [\[CrossRef\]](#)
16. Yu, X.; Sun, W.; Li, J.; Liu, K.; Liu, C.; Tan, J. ONSEP: A novel online neural-symbolic framework for event prediction based on large language model. In *Findings of the Association for Computational Linguistics ACL 2024*; Ku, L.W., Martins, A., Srikumar, V., Eds.; Association for Computational Linguistics: Bangkok, Thailand, 2024; pp. 6335–6350. Available online: <https://aclanthology.org/2024.findings-acl.378> (accessed on 23 November 2025).
17. Mathur, P.; Morariu, V.I.; Garimella, A.; Dernoncourt, F.; Gu, J.; Sawhney, R.; Nakov, P.; Manocha, D.; Jain, R. DocScript: Document-level script event prediction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino, Italy, 20–25 May 2024; Calzolari, N., Kan, M.Y., Hoste, V., Lenci, A., Sakti, S., Xue, N., Eds.; ELRA: Paris, France; ICCL: Mauren, Liechtenstein, 2024; pp. 489–498. Available online: <https://aclanthology.org/2024.lrec-main.458> (accessed on 23 November 2025).
18. Sun, Y.; Wang, S.; Li, Y.; Feng, S.; Chen, X.; Zhang, H.; Tian, X.; Zhu, D.; Tian, H.; Wu, H. Ernie: Enhanced representation through knowledge integration. *arXiv* **2019**, arXiv:1904.09223. [\[CrossRef\]](#)
19. Li, X.; Li, F.; Pan, L.; Chen, Y.; Peng, W.; Wang, Q.; Lyu, Y.; Zhu, Y. Duee: A largescale dataset for chinese event extraction in real-world scenarios. In *Proceedings of the Natural Language Processing and Chinese Computing (NLPCC 2020)*, Zhengzhou, China, 14–18 October 2020. Available online: <https://api.semanticscholar.org/CorpusID:222180086> (accessed on 23 November 2025).

20. Deng, H.; Zhang, Y.; Zhang, Y.; Ying, W.; Yu, C.; Gao, J.; Wang, W.; Bai, X.; Yang, N.; Ma, J.; et al. Title2Event: Benchmarking open event extraction with a large-scale Chinese title dataset. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; pp. 6511–6524. Available online: <https://aclanthology.org/2022.emnlp-main.437> (accessed on 23 November 2025).
21. Yao, F.; Xiao, C.; Wang, X.; Liu, Z.; Hou, L.; Tu, C.; Liu, Y.; Shen, W.; Sun, M. LEVEN: A Large-Scale Chinese Legal Event Detection Dataset. *arXiv* **2022**, arXiv:2203.08556. [\[CrossRef\]](#)
22. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
23. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Wang, S.; Hu, G. Revisiting pre-trained models for Chinese natural language processing. *arXiv* **2020**, arXiv:2004.13922v2. [\[CrossRef\]](#)
24. Guo, B.; Han, S.; Han, X.; Huang, H.; Lu, T. Label confusion learning to enhance text classification models. *arXiv* **2020**, arXiv:2012.04987. [\[CrossRef\]](#)
25. Yang, S.; Liu, L.; Xu, M. Free lunch for few-shot learning: Distribution calibration. In Proceedings of the International Conference on Learning Representations (ICLR), Vienna, Austria, 3–7 May 2021. Available online: <https://openreview.net/forum?id=JWOiYxMG92s> (accessed on 23 November 2025).
26. Yu, W.; Huang, X.; Yuan, Q.; Yi, M.; An, S.; Li, X. Information security field event detection technology based on satt-1stm. *Secur. Commun. Netw.* **2021**, 2021, 5599962. [\[CrossRef\]](#)
27. Ni, C.; Liu, W.; Li, W.; Wu, J.; Ren, H. Chinese event detection based on event ontology and siamese network. In *Knowledge Science, Engineering and Management*; Springer: Berlin/Heidelberg, Germany, 2021. Available online: <https://api.semanticscholar.org/CorpusID:237206910> (accessed on 23 November 2025).
28. Cheng, Q.; Fu, Y.; Huang, J.C.; Cheng, G.; Du, H. Event detection based on the label attention mechanism. *Int. J. Mach. Learn. Cybern.* **2022**, 14, 633–641. [\[CrossRef\]](#)
29. Yao, X.; Yang, Z.; Cui, Y.; Wang, S. Minirbt: A two-stage distilled small chinese pre-trained model. *arXiv* **2023**, arXiv:2304.00717.
30. Ke, X.; Ou, Z.; Wu, X.; Li, B. A new Chinese event detection method based on pmtnet. In Proceedings of the 16th International Conference on Machine Learning and Computing, Shenzhen, China, 2–5 February 2024. Available online: <https://api.semanticscholar.org/CorpusID:270338481> (accessed on 23 November 2025).
31. Liu, W.; Zhou, P.; Zhao, Z.; Wang, Z.; Ju, Q.; Deng, H.; Wang, P. K-BERT: Enabling Language Representation with Knowledge Graph. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), New York, NY, USA, 7–12 February 2020; pp. 2901–2908. [\[CrossRef\]](#)
32. Wang, X.; Gao, T.; Zhu, Z.; Zhang, Z.; Liu, Z.; Li, J.; Tang, J. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *Trans. Assoc. Comput. Linguist.* **2021**, 9, 176–194. [\[CrossRef\]](#)

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.