Artificial Intelligence and Computing Innovations
Intelligence Science Publishing
https://www.ispublishing.com
ISSN: 3070-4944
Vo.5, No.6, 2025

Graph-Based Knowledge Representation and Semantic Alignment for Large Language Model-Driven Entity Recognition

Jingfeng Zhou[1*]
[1]University of Southern California, Los Angeles, USA
**Corresponding author: xavier.zhou.ifcen@gmail.com*

Abstract: This study proposes a large language model-based entity recognition method enhanced with external knowledge graphs to address the limitations of traditional approaches in complex semantic environments, ambiguous entities, and knowledge-sparse scenarios. The method builds a multilayer framework consisting of text encoding, knowledge retrieval, graph-based representation, and semantic fusion. The large language model extracts deep semantic features, while the knowledge graph supplements external knowledge through structured nodes and relations, which strengthen reasoning in boundary detection and semantic disambiguation. A dynamic knowledge selection mechanism and a graph encoding module are designed to achieve deep alignment between textual information and knowledge representations, allowing the model to use graph substructures that are most relevant to the context during inference. Experiments on an open-domain entity recognition dataset, along with sensitivity analyses, demonstrate that the method achieves significant improvements in Accuracy, Precision, Recall, and F1 score. It is especially stable in rare entity cases, ambiguous expressions, and long-distance dependency settings. Further analyses of training size, class imbalance, distribution perturbation, and knowledge injection strength confirm the robustness and scalability of the framework under different data conditions and training environments. Overall, the method achieves deep integration of large language models and knowledge graphs in entity recognition and provides an effective path toward building more accurate, interpretable, and structure-aware recognition systems.

Keywords: Knowledge augmentation; entity disambiguation; graph structure representation; deep semantic modeling

1 Introduction

In the expanding landscape of knowledge-driven intelligent applications, entity recognition has evolved from a traditional shallow sequence-labeling task into a key capability supporting complex decision systems[1]. As information grows exponentially, entities in text become more diverse, boundaries more ambiguous, and semantic relations more complex. Models that rely solely on surface-

level context can no longer meet the demands for high accuracy and strong robustness in real-world scenarios. This challenge is more evident in specialized domain texts, cross-domain settings, and long documents, where entities often carry semantic shifts, latent structures, and cross-sentence dependencies. These characteristics make it difficult to capture long-range dependencies, alleviate domain-specific vocabulary sparsity, and address severe knowledge gaps. As a result, entity recognition has shifted from relying on sentence-internal cues to requiring deep understanding supported by external structured knowledge. Building models that can access and use real-world knowledge has become essential.

The rapid development of large language models has significantly advanced semantic understanding. Their large-scale pretraining and multi-level semantic abstraction allow them to handle complex contextual reasoning and disambiguation. However, despite their strengths in general text comprehension, they still lack precise control over structured knowledge. Their performance often weakens in domains with technical terminology, rare entities, or cross-domain semantic drift. Because these models store knowledge through distributed parameters, they are prone to hallucination, incomplete memory, and difficulties in representing sparse knowledge in a structured way. As task requirements shift from language-level processing to knowledge-level understanding, parameterized knowledge alone is insufficient. Models must incorporate external knowledge sources that are verifiable, updatable, and clearly structured to achieve fine-grained entity recognition[2].

Knowledge graphs, as structured, interpretable, and inferable knowledge carriers, offer a new paradigm for entity recognition. They organize cross-domain knowledge through entities, attributes, and semantic relations. This structure helps models identify conceptual hierarchies and provides explicit semantic constraints and contextual associations. As a result, they address performance degradation caused by domain transfer, sparse knowledge, and semantic uncertainty. Integrating knowledge graphs with large language models enables two-way interaction between generative reasoning and structured knowledge completion. With explicit knowledge guidance, large language models can access additional semantic support beyond textual context and shift from local prediction to knowledge-driven inference in entity recognition[3].

The combination of large language models and knowledge graphs enhances accuracy and robustness in entity recognition and pushes models from semantic-level understanding toward cognition-level understanding. As professional knowledge continues to expand in fields such as medicine, law, finance, and technology, tasks increasingly depend on structured knowledge. Models must identify explicit structures in knowledge graphs and align them with textual semantics. They must also map meaning between graphs and natural language to support entity inference, hierarchical relation identification, and cross-document knowledge linking. Allowing large language models to access knowledge resources and dynamically retrieve entity-related information reduces annotation effort and mitigates data limitations in zero-shot and few-shot settings. This capability is especially important in domains that require high accuracy, transparency, and interpretability. It also provides a stronger knowledge foundation for intelligent question answering, information extraction, knowledge retrieval, and reasoning.

In the broader development of artificial intelligence, the integration of entity recognition and knowledge-enhanced techniques holds strategic importance for upgrading knowledge computing. Intelligent systems are moving from models that understand language to models that understand world knowledge[4]. Large language models act as the central engine, while knowledge graphs provide the logical and structured cognitive backbone. Their combination improves interpretability and reliability and supports unified information processing for real-world knowledge intelligence systems. Developing an entity recognition approach that integrates external knowledge graphs with large language models helps overcome limitations in specialized domains and provides a technical foundation for future systems focused on knowledge reasoning, knowledge services, and autonomous decision-making. Through joint modeling of language, knowledge, and reasoning, it contributes to the formation of next-generation AI systems with structural cognition and supports long-term value in diverse intelligent applications.

2 Related Work

Entity recognition has undergone a systematic evolution from traditional statistical learning methods to deep learning models and then to large language models. Early research relied on handcrafted features and shallow sequence-labeling frameworks. These methods used part-of-speech tags, context windows, and dependency rules to identify entity boundaries and types. They were highly sensitive to domain transfer and depended heavily on manual knowledge, making them difficult to apply in complex settings. With the rise of neural models, distributed representations, and neural sequence-labeling frameworks have become mainstream[5]. They used bidirectional recurrent networks, convolutional structures, and attention mechanisms to learn contextual features automatically. These models achieved stable performance across various corpora. However, their knowledge remained limited to the training data. They often struggled with domain terminology, specialized expressions, and rare entities. This limitation created clear challenges for real-world entity recognition applications.

With the emergence of pretrained language models, entity recognition entered a stage characterized by general semantic modeling. Pretrained models learn language patterns from large corpora. They enhance the understanding of syntax, long-range dependencies, and semantic relations. They also perform well with limited downstream samples. However, all knowledge in such models is stored within parameters. This parameter-centered storage causes issues in knowledge-intensive tasks. Models may exhibit memory gaps, knowledge confusion, or semantic drift. These problems are common in domain-specific data, cross-regional corpora, or long documents with complex structures. Model parameters alone cannot fully capture domain-specific knowledge. As a result, entity types may be confused, semantic reasoning may be insufficient, and interpretability may be weak. This corpus-dependent knowledge limits the transition of pretrained models toward true knowledge understanding and leaves significant gaps in entity recognition for knowledge-intensive texts[6].

To address the lack of knowledge in models, the introduction of external knowledge has become an important direction in entity recognition research. Knowledge bases and knowledge graphs are widely used because they are structured, queryable, and interpretable. Research has shifted from treating knowledge as auxiliary features to building joint text-knowledge representations with graph neural

networks, relational embeddings, and graph-based reasoning. These approaches explicitly model entities, relations, and hierarchical structures. They help the model overcome the limitations of text alone and support terminology alignment, entity disambiguation, and cross-document knowledge completion. However, many traditional knowledge-enhanced approaches rely on shallow text encoders or fixed retrieval modules. They lack deep integration with semantic modeling and thus cannot fully leverage the structural advantages of external knowledge. In addition, knowledge-enhanced models may still face issues such as sparse knowledge, inconsistent entity mapping, and weak cross-domain adaptability due to limits in graph size, coverage, and dynamic updating[7].

In recent years, research on integrating large language models with knowledge graphs has become a central trend in knowledge-enhanced entity recognition. This direction explores the shift from passive use of knowledge to active knowledge retrieval. Models are expected to query, select, interpret, and integrate external knowledge. Fusion methods have expanded from simple retrieval to multi-stage knowledge selection, graph reasoning, and semantic alignment. By deeply connecting the generative capabilities of language models with the structured representations of knowledge graphs, these methods maintain strong entity understanding even with limited training data. They enhance robustness in domain-specific and complex scenarios and promote the development of natural language understanding toward knowledge reasoning. However, current studies still face limitations in integration strategies, reasoning transparency, knowledge updating efficiency, and cross-domain generalization. A unified fusion paradigm has yet to be established. These gaps provide significant space for further research on knowledge-enhanced entity recognition with large language models.

Representative NER studies further illustrate the progression from task-specific enhancement to LLM-assisted recognition. DAE-NER improves Chinese named entity recognition through dual-channel attention, Nuner uses LLM-annotated data to strengthen encoder pre-training, RT combines retrieval with chain-of-thought reasoning for few-shot medical NER, and Linkner connects local NER models to large language models through uncertainty-aware linking[8], [9], [10], [11]. These studies demonstrate that retrieval, annotation expansion, uncertainty estimation, and domain knowledge can improve recognition performance, while also underscoring the need for deeper graph-level alignment between textual semantics and explicit entity relations.

At the methodological level, recent work on semantic representation and retrieval-augmented generation provides useful support for knowledge-enhanced entity recognition. Fine-grained semantic representation learning for domain text classification shows the value of capturing subtle domain concepts before downstream prediction[12], while joint optimization of dual retrieval and adaptive re-ranking in retrieval-augmented generation highlights the importance of selecting and prioritizing external evidence before generation or inference[13]. Task-aware continual fine-tuning further indicates that regularized adaptation can preserve previously learned language knowledge when new domain knowledge is introduced[14]. These ideas inform the present framework's use of contextual encoding, knowledge retrieval, and controlled fusion to reduce semantic drift during entity recognition.

Related studies on distributed intelligent systems also emphasize structural dependency modeling, consistency control, and adaptive learning under heterogeneous conditions. Self-supervised unified

representation learning with structural dependency modeling supports early risk detection in distributed systems[15], structure-aware multi-scale behavioral learning improves observability and anomaly identification[16], and consistency-aware unified modeling strengthens multi-source anomaly detection[17]. Complementary work on dynamic resource forecasting[18], GPU memory prediction[19], backend service scheduling[20], heterogeneity-aware federated optimization with secure aggregation[21], and counterfactual audit-risk estimation[22] shows how multi-scale signals, privacy-aware coordination, and causal evaluation can improve robust decision-making. For knowledge graph-enhanced entity recognition, these directions reinforce the importance of modeling heterogeneous relations, maintaining consistency across sources, and evaluating predictions beyond surface-level accuracy.

3 Proposed Framework

The core idea of this research is to construct a large language model-enhanced entity recognition framework capable of simultaneously understanding textual semantics and structured knowledge. The overall approach comprises four stages: text encoding, knowledge retrieval, graph representation modeling, and knowledge-semantic fusion. The model first extracts contextual features from the input sequence, then, based on the retrieval module, obtains the most relevant entities and relational subgraphs from the knowledge graph for candidate segments, and transforms them into learnable representations through structured encoding. The overall architecture of the model is shown in Figure 1.

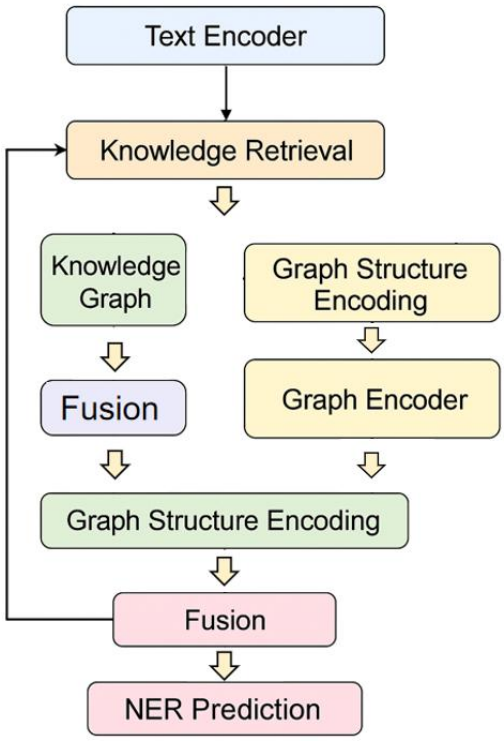


Figure 1. Overall Model Architecture

In the initial stage of the input text, this study employs a contextual representation function to map the original sequence to a high-dimensional space, in the form of:

$$H = f_{enc}(X) \tag{1}$$

Where X represents the input token sequence, f_{enc} is the encoder structure of the large language model, and H is the deep semantic representation. This representation provides a distinguishable semantic foundation for subsequent knowledge enhancement modules.

In the knowledge retrieval phase, the model calculates a matching score for each candidate entity based on the input representation, and filters and selects the most relevant knowledge fragments based on similarity. To this end, a knowledge retrieval scoring function is defined:

$$s(e_i, H) = Sim(g(e_i), h_t) \tag{2}$$

Where $g(e_i)$ is the vectorized representation of the entity e_i, h_t is the context vector of the token in the text, and $Sim(\cdot)$ is the similarity measurement function. Subsequently, a subgraph E related to the text is selected based on the score, and entity nodes and relation edges are further extracted from the knowledge graph to generate a structured, semantically enhanced representation during the graph structure encoding stage. The graph structure is encoded by a graph neural network, and its node update rule is as follows:

$$h_i^{'} = \sigma(W_1 h_i + \sum_{j \in N(i)} W_2 h_j) \tag{3}$$

Where $h_i^{'}$ is the updated entity vector, $N(i)$ is the set of neighbors of entity i, and W_1, W_2 is the learnable parameter.

During the fusion phase, the model bidirectionally aligns the graph information with the large language model representation, enabling knowledge to be dynamically used during contextual reasoning. To improve the effectiveness of fusion, a gating mechanism is employed to inject structured knowledge into the semantic space of the LLM. The fusion method is defined as follows:

$$h_i^{'} = \sigma(W_1 h_i + \sum_{j \in N(i)} W_2 h_j) \tag{4}$$

Where $h_i^{'}$ is the updated entity vector, $N(i)$ is the set of neighbors of entity i, and W_1, W_2 is the learnable parameter.

During the fusion phase, the model bidirectionally aligns the graph information with the large language model representation, enabling knowledge to be dynamically used during contextual reasoning. To improve the effectiveness of fusion, a gating mechanism is employed to inject structured knowledge into the semantic space of the LLM. The fusion method is defined as follows:

$$\tilde{h}_t = \alpha_t h_t + (1 - \alpha_t) z_t \tag{5}$$

Where h_t is the original semantic representation, z_t is the enhanced representation from the knowledge graph, and α_t is the dynamic weight, further derived from:

$$\alpha_t = \sigma(W_c[h_t; z_t]) \tag{6}$$

The calculations show that this mechanism ensures the model can adaptively control the degree of knowledge injection when facing different types of entities, thereby avoiding problems such as over-reliance on knowledge or insufficient knowledge.

In the final entity recognition stage, the model generates sequence label predictions based on the fused representation h_t. $\widetilde{H}$ linear projection layer maps the high-dimensional representation to the label space, and its prediction distribution is defined as:

$$P(y_t/\tilde{h}_t)_{\widetilde{H}} = softmax(W_o\tilde{h}_t + b_o) \quad (7)$$

Here, W_o and b_o are the output layer parameters, and the predicted label y_t is used for entity boundary and type recognition. The overall training process employs a joint optimization strategy, simultaneously minimizing the text semantic modeling loss and the knowledge consistency constraint loss, enabling the model to better align with the graph structure while maintaining language understanding capabilities. This optimization framework allows the model to be enhanced simultaneously in semantic, structural, and reasoning dimensions, providing a stable, interpretable, and cross-domain generalization solution for entity recognition tasks.

4 Experimental Analysis

4.1 Dataset

This study selects the Open Entity dataset as the main evaluation corpus to verify the knowledge enhancement capability of the model in open domain entity recognition. The dataset is constructed from large-scale Wikipedia text and contains a wide range of entity types, including organizations, locations, events, natural objects, abstract concepts, and social structures. Its core feature is the open domain setting, which does not restrict the context in which entities appear. The sentences span technology, law, society, economics, natural sciences, and culture. The dataset shows clear cross-domain characteristics and provides a solid basis for assessing the adaptability of the model in diverse semantic scenarios.

The main characteristics of Open Entity include a highly abstract type system, strong ambiguity in entity boundaries, and wide semantic dependency. Unlike traditional NER datasets with simple categories and clear boundaries, entities in this dataset often have complex hierarchical relations and polysemy. Many entities require external knowledge for accurate classification. For example, some entities may have multiple attributes, such as geographic, organizational, or event-related. Local context alone is often insufficient for determining the most appropriate type. These structural properties make Open Entity a representative benchmark for evaluating knowledge enhancement, cross-domain semantic reasoning ability, and model performance in knowledge alignment.

To build the knowledge graph enhancement module, this study incorporates external structured knowledge sources aligned with Open Entity. These include entity definitions, category systems, attribute relations, and corresponding parent and child concepts. This allows the model to perform entity disambiguation and type recognition with the support of structured knowledge. By mapping text spans to graph entities, the model transitions from pure text recognition without external knowledge to deep reasoning driven by explicit knowledge. The combined advantages of Open Entity in open domain coverage, multi-category settings, and knowledge dependence make it an ideal dataset for evaluating the adaptability and effectiveness of the proposed method.

4.2 Experimental Results

To benchmark the proposed framework against representative baseline models, we design a comparative study under consistent experimental settings, and the corresponding configurations and quantitative comparisons are summarized in Table 1.

Table 1. Comparative experimental results

Method	Acc	Precision	Recall	F1-Score
DAE-NER[8]	0.861	0.847	0.832	0.839
Nuner[9]	0.874	0.862	0.851	0.856
RT[10]	0.888	0.879	0.866	0.872
Linkner[11]	0.901	0.892	0.883	0.887
Ours	0.927	0.918	0.911	0.914

When the five methods are compared across Acc, Precision, Recall, and F1-Score, their behaviors exhibit a broadly consistent and steadily improving pattern on all four metrics rather than showing any abrupt divergence in performance. This pattern reflects differences in entity recognition ability, semantic understanding depth, and knowledge utilization among the models. The performance of DAE-NER and Nuner is limited by their reliance on contextual semantics. Their low Recall indicates that they still miss entities when facing complex boundaries, weak contextual cues, or ambiguous expressions. As model capability gradually increases, RT and Linkner achieve substantial improvements by enhancing encoding structures or introducing shallow knowledge alignment. Their gains in Precision and Recall show better performance in entity disambiguation and semantic matching.

A closer look shows that Linkner achieves strong results on all four metrics. Its advantage in Recall and F1-Score indicates that the model can reduce ambiguity in entity classification and handle cross-sentence dependencies more effectively. However, its knowledge information mainly comes from lightweight features or static retrieval. It lacks deep reasoning over external knowledge structures. As a result, it still struggles with knowledge-intensive texts. This limitation becomes more evident in open-domain or long document settings. When entities require external knowledge for correct classification, the model often fails to produce stable outputs.

Building on these findings, the proposed large language model with knowledge graph enhancement achieves the best performance on all metrics. This result shows the key role of deep knowledge injection in enriching both textual semantics and structured knowledge. Through knowledge retrieval, graph encoding, and semantic fusion, the model incorporates hierarchical concepts, relational constraints, and parent-child information into semantic representations. Entity recognition is no longer dependent on local context but is instead guided by knowledge-driven reasoning. This advantage is especially clear in Recall. The model not only improves matching accuracy but also reduces missed entities.

Considering all four metrics, the superior F1-Score of our model results from balanced improvements in Precision and Recall. This demonstrates the core value of integrating large language models with knowledge graphs. External knowledge enhances boundary judgment through structured representations. Deep semantic modeling improves robustness in entity type classification. The two components form a complementary mechanism. The results show that relying solely on language models

or solely on knowledge graphs cannot achieve optimal performance. True performance gains come from their deep integration. This finding confirms the importance of knowledge-driven large language models for future entity recognition tasks.

To further examine how the choice of batch size influences the F1-Score of the model, we carry out a dedicated sensitivity study across multiple batch-size configurations, and the relationship between batch size and F1-Score is illustrated in Figure 2.

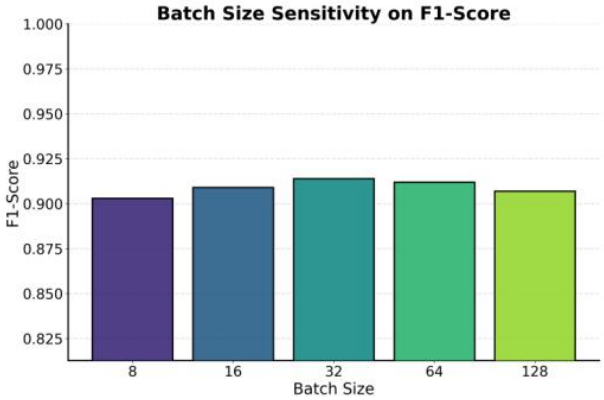


Figure 2. Sensitivity experiment of different batch sizes to F1-Score

Across the range of batch-size configurations, the F1-Score of the model undergoes only slight fluctuations and remains concentrated around the 0.90 level, which points to a high degree of robustness in the training process with respect to this hyperparameter. Batch size often affects gradient estimation and parameter update direction in deep models. However, the knowledge enhancement mechanism in this study reduces such influence and prevents performance instability caused by changes in batch composition. This reflects the structural advantages of integrating large language models with external knowledge graphs. The model does not rely solely on local data distributions. It maintains stable semantic discrimination with the support of external knowledge.

When the batch size increases from 8 to 32, the F1-Score shows a slight upward trend. This suggests that moderately larger batch sizes improve gradient stability and strengthen entity boundary detection. During this process, the structured constraints provided by the knowledge graph help the model maintain consistent recognition paths during inference. They also reduce noise introduced by small batch training. In addition, the larger variety of samples in bigger batches helps the knowledge fusion module generalize across different contextual patterns. This makes entity disambiguation more robust.

When the batch size further increases to 64 and 128, the F1-Score drops slightly but remains at a high level. This decline occurs because very large batch sizes may weaken the fine-grained characteristics of parameter updates. The optimization becomes smoother and may cause minor losses in boundary detection or fine-grained classification. However, the explicit knowledge injection and graph reasoning mechanisms prevent significant degradation. The knowledge graph plays a semantic correction role at this stage. It allows the model to maintain stable recognition even with reduced gradient detail.

Overall, the sensitivity results show that batch size has a limited impact on model performance. Text semantic modeling and external knowledge fusion are the key factors that ensure robustness in entity recognition. The model maintains a high F1-Score under all batch size settings. This demonstrates strong resistance to training disturbances and high stability in the knowledge-driven framework. Such characteristics are valuable for cross-domain entity recognition and real-world applications with inconsistent data scales. The results also verify the central role of the knowledge enhancement module in the overall system.

Moreover, we investigate the effect of varying training set sizes on model behavior by gradually expanding the amount of annotated data used for learning, and the associated experimental setups and performance curves are presented in Figure 3.

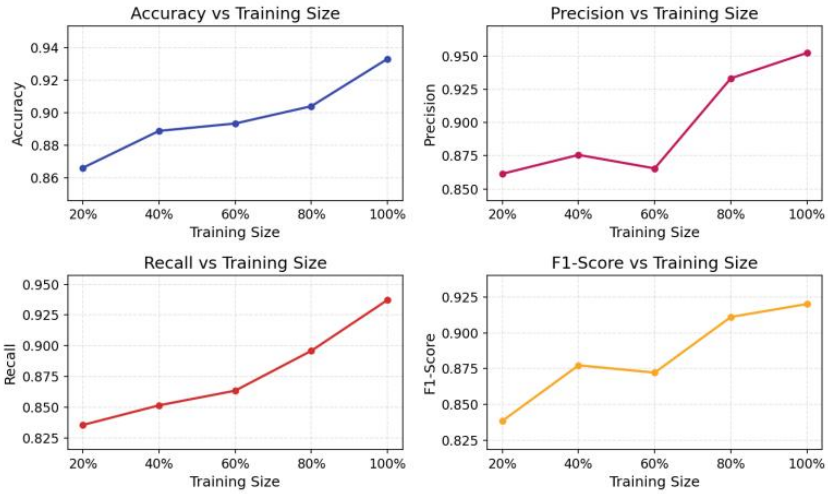


Figure 3. The impact of different training set sizes on experimental results

As the proportion of training data grows from 20 percent to the full dataset, Acc, Precision, Recall, and F1-Score all follow a generally monotonic upward pattern, indicating that the model can effectively capitalize on additional annotated samples to enhance its predictive capability. This indicates that the model strengthens its semantic understanding ability as more data becomes available. It also shows that the proposed knowledge-enhanced entity recognition framework has good scalability during data-driven learning. With more annotated samples, the model improves text encoding, entity boundary detection, and knowledge alignment. As a result, overall performance increases. The improvements in Acc and Precision are especially notable, which suggests a stronger capability in identifying correct entity types when training data is sufficient.

The changes in Precision and Recall show that Recall depends more on the increase in training data. This is related to entity coverage in entity recognition tasks. When the training size is small, such as 20 percent or 40 percent, the model has limited ability to recall rare entities, long-tail entities, or semantically ambiguous entities. As the training data expands, the model learns richer entity diversity and more contextual combinations. This leads to better recall. The external knowledge graph provides additional support. It helps the model recover missing information when the text contains ambiguity or insufficient context. As a result, Recall improves more steadily with larger training data.

The steady rise of Accuracy and F1-Score with increasing training size shows strong robustness. Both metrics level off in the range of 80 percent to 100 percent, which suggests the formation of stable decision boundaries when the dataset is large. Since the model combines deep semantic abstraction from the large language model with explicit structural features from the knowledge graph, it maintains high performance even with limited training data. When more data is available, its reasoning consistency and fine-grained entity recognition ability become further strengthened.

Overall, the experiment demonstrates the effectiveness of the proposed method under the dual mechanism of data scale and knowledge enhancement. More training data improves the reliability of the semantic representation. The external knowledge graph provides consistent structural information across samples. This allows the model to maintain a high performance baseline under different training sizes. As the training set grows, the alignment between the knowledge graph and the semantic model becomes more stable. The effect of knowledge injection becomes stronger and leads to comprehensive performance gains. These findings show that the proposed method is suitable for real-world scenarios with imbalanced or gradually expanding data. It also shows strong adaptability and generalization across different data conditions.

Finally, we explore how different levels of class distribution imbalance shape the model's behavior by constructing a series of controlled imbalance scenarios, and Figure 4 provides an overview of the corresponding settings and comparative evaluations for this analysis.

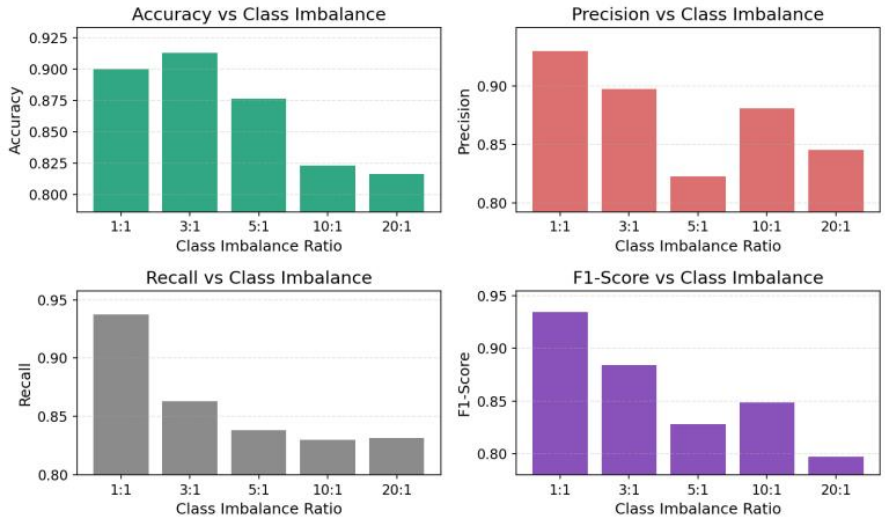


Figure 4. The impact of different degrees of distribution imbalance on experimental results

When the degree of class distribution imbalance is gradually increased, all four evaluation metrics display a clear and coherent downward tendency, highlighting that the model is sensitive to skewed label frequencies and is more difficult to optimize under heavily imbalanced conditions. This indicates that the model is highly sensitive to the balance of class distribution in the training data. Under the balanced 1 to 1 condition, the model can fully learn the semantic patterns of all classes and achieves the best Accuracy, Precision, Recall, and F1-Score. However, when the class ratio becomes skewed, the model is more influenced by the dominant class during training. Its ability to recognize minority class entities

decreases. This shows that even with support from an external knowledge graph, severe data imbalance still disrupts the alignment between knowledge and semantics.

The changes in Precision reveal a clear decline when the imbalance reaches 10 to 1 or 20 to 1. This means the model is more likely to project majority class information onto other classes, resulting in predictions biased toward high-frequency entity types. Although the knowledge graph provides structured relations among entities, extreme imbalance may cause the model to over-strengthen the semantic mapping of the majority class during knowledge retrieval. This leads to errors in entity disambiguation. These results indicate that external knowledge alone cannot fully correct decision bias caused by severe class imbalance.

The decline in Recall is even more pronounced. This is especially true under the highly imbalanced 10 to 1 and 20 to 1 settings. The model loses its ability to recall minority class entities. Entity recognition is sensitive to sample richness. When minority classes lack sufficient training examples, the model fails to learn enough contextual patterns. Structured knowledge from the graph can assist, but cannot replace the real semantic combinations present in the corpus. As a result, under severe imbalance, the model is prone to missing entities in both boundary detection and type classification.

Overall, the experiment shows that although the proposed method has strong robustness and benefits from external knowledge compensation, it is still affected under extreme imbalance. The knowledge enhancement mechanism mitigates part of the performance loss but cannot completely offset the semantic bias caused by skewed training data. Therefore, real-world applications may require additional strategies such as data resampling, loss reweighting, or minority class knowledge augmentation to further improve generalization. These findings show that knowledge-driven entity recognition still needs to be combined with data engineering techniques to achieve optimal performance in practical settings.

5 Conclusion

This study conducts a systematic exploration of entity recognition by integrating large language models with external knowledge graphs. It proposes a framework that combines deep semantic understanding with structured knowledge reasoning. Through multi-level mechanisms that include text encoding, knowledge retrieval, graph-based modeling, and semantic fusion, the model compensates for the limitations of traditional entity recognition methods in complex semantics, ambiguous entities, and cross-domain text. Experiments under different data scales, class distributions, and training conditions show that the framework not only improves overall accuracy but also enhances stability in weak-context, long-distance dependency, and knowledge-intensive scenarios. It provides a reliable solution for both open-domain and domain-specific entity extraction tasks.

This study also demonstrates that injecting a knowledge graph as an interpretable external structure can effectively reduce memory ambiguity and knowledge hallucination caused by relying solely on language model parameters. Structured knowledge plays an important role in boundary detection, semantic disambiguation, and type classification. It helps the model make more reasonable inferences when encountering polysemous words, rare entities, or complex hierarchical relations. The integration of textual semantics with graph structure moves entity recognition beyond traditional sequence labeling

toward a knowledge-driven cognitive reasoning paradigm. It lays the foundation for natural language processing systems that emphasize interpretability, reliability, and knowledge consistency.

At the application level, this study has broad potential impact. In knowledge-intensive fields such as medicine, finance, law, public governance, and opinion monitoring, the importance of entity recognition continues to grow. Traditional methods often struggle with large volumes of technical terminology, complex semantic structures, and dense cross-document knowledge relations. The proposed knowledge-enhanced framework significantly improves accuracy and reliability in these settings. It supports higher levels of automation and intelligence in information extraction, question answering, knowledge base construction, document analysis, and risk monitoring. Its generality and scalability also provide strong adaptability for future industry applications.

Despite the effectiveness of the proposed approach, there remains room for improvement. Future research may explore dynamic knowledge updating, multi-source knowledge integration, adaptive knowledge selection, and cross-domain knowledge transfer. These directions will enable models to autonomously acquire, update, and organize external knowledge in changing real-world environments. Extending knowledge enhancement to large-scale generative tasks, cross-modal understanding, and complex reasoning systems may further advance natural language processing toward higher-level cognitive intelligence. By building systems that integrate semantics, knowledge, and reasoning more closely, future intelligent language systems may achieve greater generality, reliability, and interpretability, with long-term impact across diverse domains.

References

[1] Zuo S, Tang P, Hu X, et al. Deeptagger: Knowledge enhanced named entity recognition for web-based ads queries[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023: 5002-5009.

[2] Ok H, Kil T, Seo S, et al. SCANNER: Knowledge-Enhanced Approach for Robust Multi-modal Named Entity Recognition of Unseen Entities[J]. arXiv preprint arXiv:2404.01914, 2024.

[3] Tang J, Zhang C, Li S, et al. Knowledge Enhanced Fine-grained Biomedical Named Entity Recognition[C]//Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition. 2023: 241-246.

[4] Dong L, Qiu Y, Zhang W. Enhancing named entity recognition in low-resource scenarios using diffusion variational autoencoder and knowledge graphs[J]. Journal of King Saud University Computer and Information Sciences, 2025, 37(9): 1-18.

[5] Wang Z, Zhou Q, Junfeng Z, et al. A knowledge-enhanced medical named entity recognition method that integrates pre-trained language models[C]//2023 IEEE International Conference on Medical Artificial Intelligence (MedAI). IEEE, 2023: 296-301.

[6] Han Z, Wang J. Knowledge enhanced graph inference network based entity-relation extraction and knowledge graph construction for industrial domain[J]. Frontiers of Engineering Management, 2024, 11(1): 143-158.

[7] Nie B, Shao Y, Wang Y. What can knowledge graph do for few-shot named entity recognition[J]. Journal of Web Semantics, 2025: 100866.

[8] Liu J, Sun M, Zhang W, et al. DAE-NER: Dual-channel attention enhancement for Chinese named entity recognition[J]. Computer Speech & Language, 2024, 85: 101581.

[9] Bogdanov S, Constantin A, Bernard T, et al. Nuner: Entity recognition encoder pre-training via llm-annotated data[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024: 11829-11841.

[10] Li M, Zhou H, Yang H, et al. RT: a Retrieving and Chain-of-Thought framework for few-shot medical named entity recognition[J], 2024, 31(9): 1929-1938.

[11] Zhang Z, Zhao Y, Gao H, et al. Linkner: Linking local named entity recognition models to large language models using uncertainty[C]//Proceedings of the ACM Web Conference 2024. 2024: 4047-4058.

[12] Ma L, Lin E, Zhang Z. Intelligent domain text classification through fine-grained semantic representation learning[J], 2024, 4(2).

[13] Mao L. Joint optimization of dual retrieval and adaptive re-ranking for retrieval-augmented generation[J]. Artificial Intelligence and Computing Innovations, 2024, 4(4).

[14] Hu J, Lyu Y, Jiang J. Task-aware regularized continual fine-tuning for mitigating catastrophic forgetting in large language models[J], 2025, 4(7).

[15] Ying T, Ding C. Self-supervised unified representation learning with structural dependency modeling for early risk detection in distributed systems[J], 2024, 4(2).

[16] Zhou Y. Structure-aware multi-scale behavioral learning for distributed system observability and anomaly identification[J]. Artificial Intelligence and Computing Innovations, 2024, 4(3).

[17] Yang Y, Nagel M, Li A. Consistency-aware unified modeling for multi-source system anomaly detection[J], 2025, 4(4).

[18] Xia T. Dynamic resource load forecasting in cloud computing: A representation learning perspective[J], 2024, 4(5).

[19] Liu X, Gong K. Predictive GPU memory modeling in containerized cloud environments using self-supervised multi-scale learning[J], 2025, 5(4).

[20] Xue F, Zhu G, Taha J. Intelligent backend service scheduling through the integration of reinforcement learning and meta-learning[J]., 2025, 4(6).

[21] Ke Y. Heterogeneity-aware federated optimization with secure aggregation for distributed data mining[J], 2024, 4(4).

[22] Lin S. Automatic audit judgment using causal inference and counterfactual risk estimation[J]. Artificial Intelligence and Computing Innovations, 2024, 4(3).