AnchorAlign: A Novel Anchor Alignment-enhanced Generative Method for Joint Named Entity Recognition and Relation Extraction

Xiaolong Weng[1], Yuanyun Zhou[1], Boyu Qiu[1], Zehua Wang[1]
Ying Xiong[1]*, Buzhou Tang[1,2]*
[1]Harbin Institute of Technology, Shenzhen, China
[2]Pengcheng Laboratory, Shenzhen, China
tangbuzhou@gmail.com

Abstract

Named Entity Recognition (NER) and Relation Extraction (RE) are two fundamental and interdependent tasks in information extraction (IE), aiming to identify entities and relations from unstructured text. Recently, generative methods have become mainstream instead of discriminative methods for IE, especially joint multi-task IE, due to their promising performance and flexibility. For joint NER and RE, existing methods suffer from misalignment between entities and relations, as well as misalignment among relations. To address these issues, we propose AnchorAlign, a novel generative method enhanced by anchor alignment. Specifically, we first introduce an anchor entity selection mechanism to identify key entities in the text as anchor points, which serve as semantic pivots to bridge the two tasks. Then, we design a dual-level anchor alignment module: at the semantic level, we construct a cross-task semantic alignment space to align the semantic representations of anchor entities and their associated relations; at the generation level, we introduce an anchor-guided generation constraint to guide the model to generate entities and relations with strict alignment based on the anchor points. Extensive experiments on five benchmark datasets show that AnchorAlign outperforms SOTA baselines, demonstrating its effectiveness. Our work provides a new perspective for optimizing the joint modeling of NER and RE, and has potential to be extended to more complex multi-task IE such as NER and Event Extraction (EE).

1 Introduction

Information Extraction (IE) is a fundamental technology for converting unstructured text into structured knowledge. Traditional approaches decompose IE into isolated sub-tasks such as Named Entity Recognition (NER), Relation Extraction (RE), and Event Extraction (EE), while recent advancements have shifted to joint extraction frameworks, particularly generative methods. These joint multi-task IE frameworks based on generative methods, including Universal Information Extraction (UIE) (Lu et al., 2022; Lou et al., 2023; Wang et al., 2023; Xiao et al., 2023; Li et al., 2024; Sainz et al., 2023), have demonstrated promising performance with high flexibility.

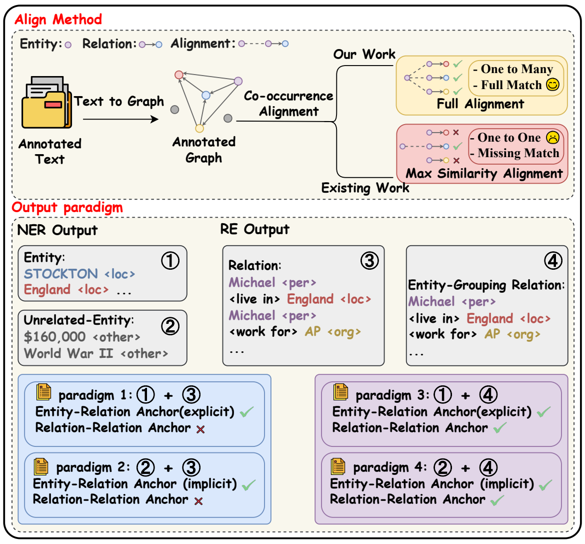

Figure 1: Comparative Analysis of Structured Output Paradigms and Alignment Mechanisms in generative methods for joint NER and RE.

In the case of typical IE tasks, NER and RE, existing joint generative methods still have the following limitations in structured output paradigms and multi-task alignment.

First, although some effective structured output paradigms (e.g., paradigm 1 (Asada and Miwa, 2025) and 2 (Lu et al., 2022) in Figure 1) have been proposed for generative models, there is no study to compare and unify them.

Second, the alignment mechanisms in most existing generative methods only focus on inter-task dependencies (i.e., dependencies between NER and

*Corresponding authors.

Findings of the Association for Computational Linguistics: ACL 2026, pages 38068–38081
July 2-7, 2026 ©2026 Association for Computational Linguistics

RE), ignoring intra-task dependencies (e.g., dependencies among entity relations). In the case of the alignment mechanisms for dependencies between NER and RE, auxiliary objectives are specially designed to align entities in NER with entities in RE. The existing alignment mechanisms (Zheng et al., 2021; He and Tang, 2023) usually align one entity in NER to the closet entity in one relation in RE no matter how many relation the same entity in. They may lead to some biases as shown in Figure 1.

To address the limitations mentioned above, we propose AnchorAlign, a novel generative method enhanced by anchor alignment with the following advantages:

- We introduce anchor entity to unify the existing output paradigms of generative methods, and two new paradigms (i.e., paradigms 3 and 4 in Figure 1) are derived them in the unified way.
- Based on anchor entities, we design new semantic alignment mechanism, which not only considers intra-task (RE) dependencies, but also avoids biases caused by inter-task (NER and RE) dependencies.
- Extensive experiments on five benchmark datasets demonstrate the effectiveness of our proposed method, outperforming other state-of-the-art (SOTA) methods for comparison. The introduced new output paradigms as well as semantic alignment mechanism bring significant improvement.

2 Related Work

2.1 Generative Information Extraction

Traditional IE typically relies on discriminative architectures to classify text spans into predefined schemas, often necessitating complex, task-specific engineering. In contrast, generative IE flattens structured information into linear sequences for direct output generation.

Generative paradigms have recently eclipsed discriminative methods to become the dominant framework for IE. The landscape of generative IE is largely defined by three primary paradigms for formulating the output, differentiated by how they represent the target structured information:

- **Structured Language Generation** This paradigm represents the inherent structure of the information by using special symbolic tokens or by introducing new, semantically initialized token embeddings. This approach is exemplified by models such as BART-NER (Yan et al., 2021), REBEL (Cabot and Navigli, 2021), UIE (Lu et al., 2022), and USM (Lou et al., 2023).
- **Natural Language Generation** This paradigm represents the target structured information as a natural language sequence. It typically leverages prompts, often including extraction guidelines or annotations, to guide a general-purpose Language Model toward the correct task interpretation. This method is employed by a wide range of models, including TANL (Paolini et al., 2021), DEEPSTRUCT (Wang et al., 2022), GenIE (Josifoski et al., 2022), and InstructUIE (Wang et al., 2023).
- **Code Style Generation** This paradigm is distinguished by its use of a Code-LLM as the base model. It reframes the extraction task by converting the schema into code constructs (e.g., classes) and represents the structured output using code-style syntax. Notable examples in this category are CODEIE (Li et al., 2023), GoLLIE (Sainz et al., 2023), and KnowCoder (Li et al., 2024).

Our work builds upon the generative architecture of BARTNER (Yan et al., 2021), extending it with a structured generation language designed for both NER and RE, thereby situating our approach within the paradigm of structured language generation.

2.2 Joint Extraction

Joint extraction aims to unify multiple IE sub-tasks within a single model. However, realizing in discriminative architectures remains difficult due to the rigid coupling of task-specific modules. Designing specialized decoding structures for heterogeneous tasks often leads to combinatorial complexity and conflicting optimization objectives, making it challenging to extend these models beyond basic tasks. Consequently, discriminative joint extraction research (Ye et al., 2022; He and Tang, 2023) has predominantly converged on NER and RE, exploiting the strong correlation where NER acts as a natural and tractable precursor to RE.

Generative architectures, conversely, offer a more flexible paradigm by unifying any IE sub-

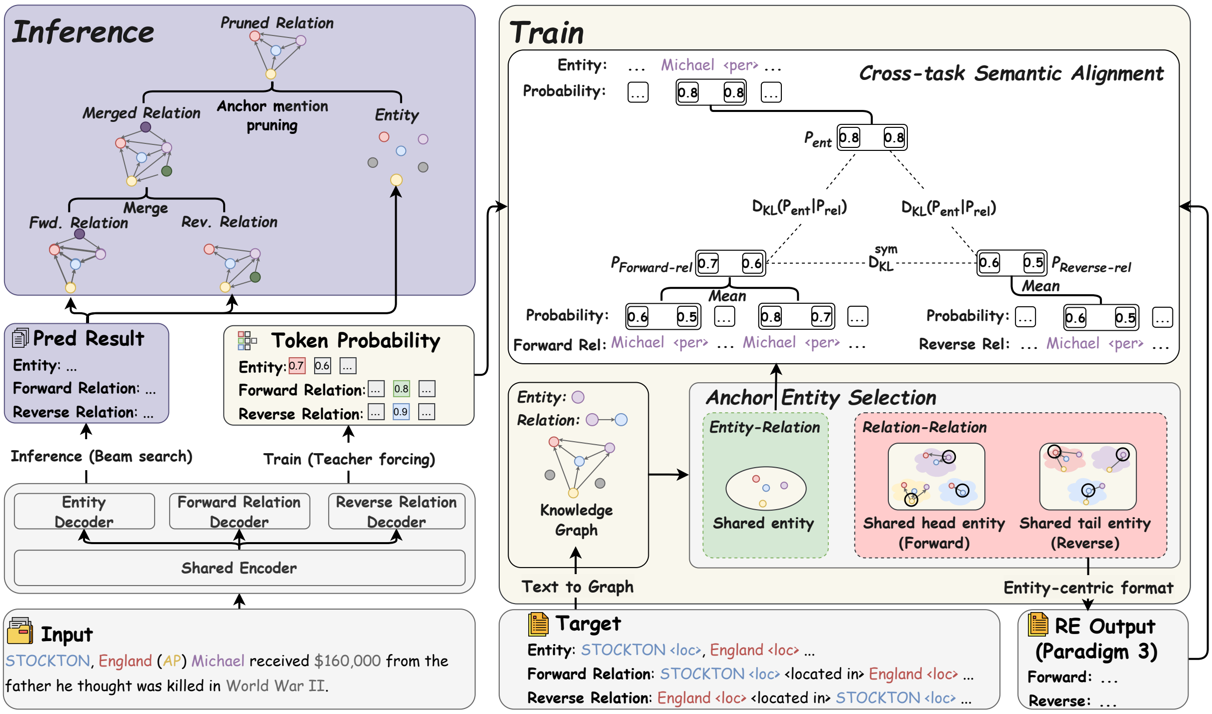

Figure 2: Overview of the AnchorAlign Framework

task into a text generation problem. This inherent flexibility significantly lowers the barrier to integrating diverse sub-tasks, which has led to the development of numerous Universal Information Extraction (UIE) models capable of concurrently handling NER, RE, and EE, such as UIE (Lu et al., 2022), InstructUIE (Wang et al., 2023), and KnowCoder (Li et al., 2024). Our work is situated within the generative paradigm for joint NER and RE. It yields significant performance gains on the relation extraction task, driven by a strategic enhancement of inter-task consistency.

3 Methodology

This section details the methodology of AnchorAlign. We first describe the Anchor Entity Selection Mechanism, followed by the model's output paradigm. Finally, we introduce a dual-level anchor alignment mechanism, comprising cross-task semantic alignment for training and anchor-guided generation constraints for inference. These components are integrated into a unified framework. The overall architecture is illustrated in Figure 2.

3.1 Anchor Entity Selection Mechanism

Alignment in joint extraction encompasses two distinct scenarios: entity-relation alignment and relation-relation alignment.

Both alignment types are facilitated by key entities that link disparate task components. We define the entities shared between NER and RE as **anchors for entity-relation**, while entities shared among multiple relations function as **anchors for relation-relation**.

In the following sections, we elaborate on the detailed designs of the output paradigm, training objectives, and inference strategies, all of which are centered around the aforementioned anchors.

3.2 Output Paradigm

AnchorAlign represents labels as unique special tokens (e.g., <PER>, <WORK FOR>) integrated into the vocabulary. To leverage semantic priors, these tokens are initialized by mean-pooling the embeddings of their original label text:

$$\mathrm{E}_{\mathrm{label}} = \mathrm{Mean}\big(\mathrm{TokenEmbed}(\mathrm{text}_{\mathrm{label}})\big) \quad (1)$$

This single-token format eliminates boundary markers (e.g., "()") and shortens sequences, which are then generated by two specialized decoders:

NER Decoder This decoder generates a linearized sequence of entities, where each entity (e_i) is represented as a pair of its textual mention (m_i) and its corresponding single-token type ($\langle t_i \rangle$):

$$\mathrm{O}_{\mathrm{NER}} = (\underbrace{m_1, \langle t_1 \rangle}_{e_1}, \ldots, \underbrace{m_n, \langle t_n \rangle}_{e_n}) \quad (2)$$

RE Decoder To align **anchors for relation-relation**, this decoder employs a subject-centric, factorized format. All relations for a given subject entity (e_{s_i}) are grouped into a single block. The subject is generated once, followed by all its corresponding object-relation pairs ($e_{o_{i,j}}, \langle r_{i,j} \rangle$):

$$\mathrm{O}_{\mathrm{RE}} = (\underbrace{e_{s_1}, e_{o_{1,1}}, \langle r_{1,1} \rangle, e_{o_{1,2}}, \langle r_{1,2} \rangle,}_{\text{Block for } e_{s_1}} \quad \underbrace{e_{s_2}, \dots}_{\text{Block for } e_{s_{2\sim n}}} \quad) \tag{3}$$

Similarly, the reverse RE decoder adopts an object-centric strategy.

This output paradigm simultaneously optimizes generation efficiency and ensures the alignment of **anchors for relation-relation**.

3.3 Cross-task Semantic Alignment

3.3.1 Positive Alignment

Our core design principle is to enforce alignment within the output probability space rather than the internal feature space. This strategy ensures consistency at the decision level while accommodating inherent disparities in internal representation capacities across different task components.

Positive Sampling Alignment comprises two components: NER-RE alignment and Forward-Reverse RE alignment. Both utilize Kullback-Leibler divergence (D_{KL}) as the objective.

NER-RE Alignment To facilitate the semantic alignment of **anchors for entity-relation**, we align both the forward (d = fwd-RE) and reverse (d = rev-RE) RE decoders with the NER output in the probability space. We define the alignment anchors as the co-occurring entity components (i.e., entity mentions and types) shared between the NER and RE tasks.

Let $p_{\mathrm{NER}}(\cdot|e')$ and $p_d(\cdot|e)$ denote the predictive probability distributions defined over the restricted vocabulary $\mathcal{V}_{\text{restricted}}$, which represent the token-level generation probabilities for the shared anchor entities in the NER and RE outputs, respectively. Treating the NER output as a high-confidence "teacher" signal, we constrain the RE predictive distribution to align with the NER target distribution via Kullback-Leibler (KL) divergence:

$$\mathcal{L}_{\mathrm{NER},d} = \frac{1}{|\mathcal{E}_d|} \sum_{e \in \mathcal{E}_d} D_{\mathrm{KL}} \left(p_{\mathrm{NER}}(\cdot|e') \| p_d(\cdot|e) \right) \tag{4}$$

where $\mathcal{E}_d$ is the set of anchor entities from the target sequence of decoder d, and $e' \in \mathcal{E}_{\mathrm{NER}}$ is the corresponding anchor entity in the NER output.

The final NER-RE alignment loss is the average of these two components:

$$\mathcal{L}_{\text{NER-RE}} = \frac{1}{2} \left(\mathcal{L}_{\text{NER, fwd-RE}} + \mathcal{L}_{\text{NER, rev-RE}} \right) \tag{5}$$

Forward-Reverse RE Alignment To mitigate biases inherent in the generation order and indirectly reinforce **anchors for relation-relation** alignment, we enforce consistency between the forward and reverse RE decoders. In this context, the target domains are the token sequences forming the ground-truth relation triplets τ (comprising the subject, relation type, and object) shared between both decoders. We designate these shared triplets as our alignment anchors.

We define $p_{\text{fwd}}(\tau)$ and $p_{\text{rev}}(\tau)$ as the predictive probability distributions defined over the restricted vocabulary $\mathcal{V}_{\text{restricted}}$. Specifically, they represent the sequence of token-level generation probabilities for these shared anchor triplets within the forward and reverse RE outputs, respectively. To align these sequences of predictive distributions, we introduce a symmetric Kullback-Leibler (KL) divergence loss:

$$D_{\mathrm{KL}}^{\mathrm{sym}}(p_1 \| p_2) = \frac{1}{2} \left(D_{\mathrm{KL}}(p_1 \| p_2) + D_{\mathrm{KL}}(p_2 \| p_1) \right) \tag{6}$$

The final Forward-Reverse RE alignment loss is the average symmetric divergence over all ground-truth triplets $\tau \in \mathcal{T}$:

$$\mathcal{L}_{\text{RE-Align}} = \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} D_{\mathrm{KL}}^{\mathrm{sym}}(p_{\text{fwd}}(\tau) \| p_{\text{rev}}(\tau)) \tag{7}$$

3.3.2 Negative Suppression

Although positive alignment ensures consistency for positive samples, it lacks explicit penalties for valid but incorrect tokens. To address this, we introduce a negative suppression mechanism to complement the positive alignment by strictly discouraging non-target generation.

Formulation of the Negative Set Let $\mathcal{V}_{\mathrm{NER}}$ and $\mathcal{V}_{\mathrm{RE}}$ represent the ground-truth token sets. The negative set $\mathcal{V}_{\text{neg}}$ is defined as the restricted vocabulary $\mathcal{V}_{\text{restricted}}$ (comprising tokens of input, label and special symbols) excluding these targets:

$$\mathcal{V}_{\text{neg}} = \{ v \in \mathcal{V}_{\text{restricted}} \mid v \notin \mathcal{V}_{\mathrm{NER}} \wedge v \notin \mathcal{V}_{\mathrm{RE}} \} \tag{8}$$

Suppression Loss We apply the suppression objective independently to both RE decoders to minimize the probability mass assigned to the negative set $\mathcal{N}$. For a decoder $d \in \{\text{fwd-RE}, \text{rev-RE}\}$ with target sequence length T_d, the loss is formulated as:

$$\mathcal{L}^d_{\text{suppress}} = -\frac{1}{T_d}\sum_{t=1}^{T_d}\frac{1}{|\mathcal{N}|}\sum_{v\in\mathcal{N}}\log(1 - p_d(v|\mathbf{h}_t)) \tag{9}$$

where $p_d(v|\mathbf{h}_t)$ is the predicted probability of a negative token v at step t. The final suppression loss is the average over both directions:

$$\mathcal{L}_{\text{suppress}} = \frac{1}{2}\left(\mathcal{L}^{\text{fwd-RE}}_{\text{suppress}} + \mathcal{L}^{\text{rev-RE}}_{\text{suppress}}\right) \tag{10}$$

This objective effectively sharpens the predictive distribution by penalizing the generation of irrelevant tokens for the given instance.

3.3.3 Overall Training Objective

The overall training objective integrates standard generative extraction tasks with our proposed anchor alignment module. We first define the total generation loss, $\mathcal{L}_{\text{Gen}}$, as the aggregation of the cross-entropy (CE) loss $\mathcal{L}^d_{\text{CE}}$ from each decoder d in the set $\mathcal{D} = \{\text{NER}, \text{fwd-RE}, \text{rev-RE}\}$:

$$\mathcal{L}_{\text{Gen}} = \sum_{d\in\mathcal{D}}\mathcal{L}^d_{\text{CE}} \tag{11}$$

The final objective $\mathcal{L}$ combines this generation loss with the auxiliary alignment and suppression terms:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{Gen}} + \alpha\mathcal{L}_{\text{NER-RE}} \\ + \beta\mathcal{L}_{\text{RE-Align}} + \gamma\mathcal{L}_{\text{suppress}}\end{aligned} \tag{12}$$

where α, β, γ are hyperparameters weighing the NER-RE alignment, forward-reverse RE alignment, and negative suppression terms, respectively.

3.4 Anchor-guided Generation Constraint

While aggregating forward and reverse RE outputs maximizes recall, it inevitably introduces false positives. To mitigate this, we employ an inference-time **Anchor-guided constraint filter**.

Principle We designate mentions of **anchors for entity-relation** as generation constraints to guide the model to generate relations with strict alignment based on the anchor points.

Specifically, by treating the NER output as a high-confidence reference, we cross-validate the entity mentions within the predicted relation triplets. Any triplet containing an entity mention unrecognized by the NER module is classified as low-confidence and subsequently discarded.

Formulation Let $\hat{\mathcal{E}}_{\text{NER}}$, $\hat{\mathcal{T}}_{\text{fwd}}$, and $\hat{\mathcal{T}}_{\text{rev}}$ denote the prediction sets from the NER and the two RE decoders, respectively. We first form a candidate pool by aggregating the RE outputs: $\hat{\mathcal{T}}_{\text{cand}} = \hat{\mathcal{T}}_{\text{fwd}} \cup \hat{\mathcal{T}}_{\text{rev}}$. We then apply the generation constraint to derive the final set $\hat{\mathcal{T}}_{\text{final}}$:

$$\begin{aligned}\hat{\mathcal{T}}_{\text{final}} = \{(s, r, o) \in \hat{\mathcal{T}}_{\text{cand}} \mid \\ \{s_{\text{mention}}, o_{\text{mention}}\} \subseteq \hat{\mathcal{E}}_{\text{NER}}\}\end{aligned} \tag{13}$$

This constraint guarantees high precision while preserving recall gains.

3.5 Overall Training Framework

Drawing inspiration from prior works (Lu et al., 2022; Lou et al., 2023; Li et al., 2024), we adopt a two-stage training framework: task-adaptive pre-training followed by supervised fine-tuning (SFT). Specifically, the pre-training phase utilizes large-scale LLM-augmented corpora to equip the backbone with an understanding of structural paradigms and common IE capabilities.

3.5.1 Data Augmentation

Inspired by guidance-based prompting for NER (Kang et al., 2024), we devise a data augmentation strategy tailored for joint NER and RE extraction. We employ DeepSeek-V3 to generate high-quality pseudo-labeled instances by prompting it with the original training exemplars. Specific details are provided in Appendix A.

3.5.2 Pre-training Stage

Following augmentation, the base model $\mathcal{M}$ undergoes a two-step pre-training phase on the generated dataset to internalize joint extraction capabilities.

1. **Entity Type Understanding** We replace entity mentions in the augmented data with type-specific mask tokens ($\langle \text{mask}_i \rangle$). Training on this masked data forces the model to learn context-type associations rather than memorizing surface forms.

2. **Progressive Objective Activation** Training on the unmasked dataset proceeds in two sequential phases. Initially, standard NER and RE objectives are employed to secure fundamental capabilities. Once the model produces semantically valid outputs, the cross-task semantic alignment objectives are activated to explicitly inject cross-task consistency.

Upon completion, the pre-trained backbone is termd $\mathcal{M}'$.

3.5.3 Supervised Fine-Tuning (SFT) Stage

The SFT stage adapts $\mathcal{M}'$ to the downstream ground-truth dataset $\mathcal{D}_{\text{train}}$ using a curriculum learning paradigm. We partition $\mathcal{D}_{\text{train}}$ based on triplet density, defining a simple-sample subset $\mathcal{D}_{\text{simple}}$ with a triplet count threshold K:

$$\mathcal{D}_{\text{simple}} = \{x \in \mathcal{D}_{\text{train}} \mid |\mathcal{T}_x| \leq K\} \quad (14)$$

The optimization proceeds in two phases:

1. **Curriculum Warm-up.** Training initiates on $\mathcal{D}_{\text{simple}}$ using only standard generation objectives. This facilitates smooth adaptation to the target distribution's easy samples.

2. **Full Data Training.** We then transition to the complete dataset $\mathcal{D}_{\text{train}}$. In this phase, we rigorously apply the **Progressive Objective Activation** protocol (Sec. 3.5.2).

4 EXPERIMENTS

4.1 Experiment Setups

Datasets. We conduct experiments on five widely used benchmark datasets, including ACE05 (Walker and Consortium, 2005), SciERC (Luan et al., 2018), ADE (Gurulingappa et al., 2012), NYT (Riedel et al., 2010), and Text2DT (Zhu et al., 2023).

Evaluation Metrics. Following prior works, we adopt two metrics to evaluate RE performance:

1. **Relation Strict F1.** A predicted relation is correct if the mentions and types of both the subject and object entities, as well as the relation type, are all correctly identified.

2. **Relation Boundary F1.** A predicted relation is correct if the mentions of both the subject and object entities and the relation type are correctly identified, regardless of the correctness of entity types.

Hyperparameter Configuration. On each dataset, we manually tuned hyper-parameters and retained the configuration that yielded the highest relation F1 on the development set (10-fold cross-validation for ADE). Please refer to Appendix D for the detailed settings.

4.2 BASELINES

We select several SOTA models for relation extraction, which can be categorized into generative and discriminative paradigms.

Generative Models. We compare our method against the following SOTA generative baselines: **UIE** (Lu et al., 2022) unifies various IE tasks by reformulating them into a text-to-structure generation format. **REBEL** (Cabot and Navigli, 2021) treats relation extraction as a sequence-to-sequence task, fine-tuned on the BART-Large architecture. **YAYI-UIE** (Xiao et al., 2023) is an end-to-end model based on Baichuan2-13B that enhances performance by co-training on both dialogue and IE data. **InstructUIE** (Wang et al., 2023) fine-tunes FlanT5-11B using natural language instructions to guide the generation process. Finally, **KnowCoder** (Li et al., 2024), based on LLaMA2-7B, reframes IE tasks as a Python code generation problem within a two-stage fine-tuning framework.

Discriminative Models. We compare against competitive discriminative baselines: **PL-Marker** (Ye et al., 2022) models relations between spans via a packed levitated marker strategy. **BiSPN** (He and Tang, 2023) employs set prediction for parallel entity and relation extraction to mitigate error propagation. **USM** (Lou et al., 2023) decomposes IE into token-token and label-token linking subtasks within a unified pipeline.

4.3 Main Results

Table 1 summarizes the comprehensive performance comparison across five public datasets, reporting both Relation Strict F1 and Relation Boundary F1.

In terms of Relation Strict F1, AnchorAlign achieves new SOTA results on SciERC and ADE, surpassing previous bests by **0.61%** and **0.14%**, respectively. On ACE05, it remains highly competitive, ranking second only to the discriminative model USM by a narrow margin of **0.65%**.

Regarding Relation Boundary F1, our model establishes new SOTA performance on four out of five datasets, achieving substantial gains on ACE05 (**+0.72%**), SciERC (**+1.00%**), ADE (**+0.88%**), and Text2DT (**+0.12%**). Although slightly trailing the discriminative baseline USM on NYT, AnchorAlign surpasses the prior SOTA average by **0.46%** overall.

Two baseline discrepancies warrant notice: (1) PL-Marker's single-label constraint is incompatible with NYT's multi-label nature (allowing multiple relations per entity pair). (2) Second, UIE's lower score on Text2DT stems from the limited capacity of the `uie-char-small` variant compared to its

Metric	Architecture	Method	ACE05	SciERC	ADE	NYT	Text2DT	AVG
Strict	Discriminative	BiSPN (He and Tang, 2023)	63.74	36.65	83.70	–	–	61.36
		PL-Marker (Ye et al., 2022)	62.80	36.93	83.20	–	–	60.98
		USM (Lou et al., 2023)	67.88	37.36	–	–	–	–
	Generative	UIE (Lu et al., 2022)	66.06	36.53	82.10	–	–	61.56
		REBEL (Cabot and Navigli, 2021)	–	–	82.21	–	–	–
	–	SoTa	**67.88**	37.36	83.70	–	–	62.98
	Generative	**Ours**	67.23	**37.97**	**83.84**	–	–	63.01
Boundary	Discriminative	BiSPN (He and Tang, 2023)	66.29	46.40	84.42	92.56	94.90	76.91
		PL-Marker (Ye et al., 2022)	65.72	47.81	84.23	80.64	93.30	74.34
		USM (Lou et al., 2023)	–	–	–	94.07	–	–
	Generative	UIE (Lu et al., 2022)	68.58	48.30	82.10	93.40	83.31	75.14
		REBEL (Cabot and Navigli, 2021)	–	–	–	91.96	–	–
		InstructUIE (Wang et al., 2023)	–	45.15	82.31	90.47	–	–
		YAYI-UIE (Xiao et al., 2023)	–	40.94	84.14	89.97	–	–
		KnowCoder (Li et al., 2024)	64.50	40.00	84.30	93.70	–	–
	–	SoTa	68.58	48.30	84.42	**94.07**	94.90	78.05
	Generative	**Ours**	**69.30**	**49.30**	**85.30**	93.64	**95.02**	**78.51**

Table 1: Performance comparison on five public datasets. We report **Relation Strict F1** and **Relation Boundary F1**. Consistent with prior research, we evaluate NYT and Text2dt solely on Relation Boundary F1.

larger English counterparts, likely constrains its representation capacity for this task.

These results demonstrate that AnchorAlign effectively surpasses prior methods, particularly in comparison to generative peers, validating the efficacy of our structural design.

4.4 Output Paradigm Design Comparison

As illustrated in Figure 1, we propose two output configurations for NER and RE tasks, respectively:

- **NER Outputs (Circles 1 & 2):** Circle 1 generates a full concatenation of all entities, while Circle 2 provides a simplified version containing only non-relational entities.
- **RE Outputs (Circles 3 & 4):** Circle 3 outputs a concatenation of all relation triplets. Circle 4 implements an entity-centric factorization, where each unique subject/object is produced followed by all associated object/subject-relation pairs.

By combining these configurations, we derive four paradigms distinguished by anchor utilization:

- **Paradigms 1 and 2:** Paradigm 1 explicitly includes **anchors for entity-relation** within the NER output. Conversely, Paradigm 2 leverages them implicitly by decoupling them from the NER component. Notably, both paradigms lack a mechanism to leverage **anchors for relation-relation**.
- **Paradigms 3 (winner) and 4:** These extend the previous two by incorporating entity-centric factorization (Circle 4). By leveraging **anchors for relation-relation**, they effectively bridge relations that share common entities.

Detailed comparative experiments and analyses regarding the impact of shared entity density are provided in Appendix B.

4.5 Ablation Study

Table 2 presents an ablation study based on Relation Strict F1, distinguishing between training-related and inference-related components:

Training-Related Components. 1) w/o Pre-training: Removing the pre-training stage leads to a 1.17% decrease in F1. This demonstrates that pre-training effectively injects prior IE knowledge into the model and enhances its comprehension of structured output paradigms. **2) w/o All Semantic Alignment:** Excluding the complete cross-task semantic alignment objective causes the most significant overall performance drop (2.00%), underscoring its criticality. When breaking down this module, removing the NER-RE Alignment yields the largest individual decrease (1.48%), highlighting the primary importance of bridging entity and relation representations. Furthermore, omitting the Forward-Reverse RE Alignment and Negative Suppression results in notable drops of 1.17% and 1.03%, respectively, validating their complemen-

Method	Strict F1	△ F1
AnchorAlign (Full Model)	**67.23**	–
Impact of Training Components		
w/o Pre-training	66.06	-1.17
w/o All Semantic Alignment	65.23	-2.00
w/o NER-RE Alignment	65.75	-1.48
w/o Forward-Reverse RE Alignment	66.06	-1.17
w/o Negative Suppression	66.20	-1.03
Impact of Inference Strategies		
w/o Generation Constraint	65.51	-1.72
w/o Decoder Aggregation	65.57	-1.66
w/o Aggregation & Constraint	64.51	-2.72

Table 2: Ablation study of different components on the ACE05 dataset. We report the Relation Strict F1 and the performance drop (Δ) when removing each component.

Data Subset	Method	NER F1	RE Strict F1	△ RE F1
Full Test Set (551 Sentences)	Baseline	67.43	37.20	–
	AnchorAlign	**67.71**	**37.97**	**+0.77**
Nested Entities (17 Sentences)	Baseline	**69.49**	29.63	–
	AnchorAlign	67.80	**29.79**	**+0.16**
Overlapping Relations (224 Sentences)	Baseline	**71.29**	38.80	–
	AnchorAlign	70.73	**41.44**	**+2.64**

Table 3: Performance comparison on complex structures (nested entities and overlapping relations) using the SciERC dataset.

Method	Parameters	Inference Time	Throughput
BiSPN (He and Tang, 2023)	199M	24.3 ms/sample	41.21 samples/s
PL-Marker (Ye et al., 2022)	219M	47.8 ms/sample	21.92 samples/s
AnchorAlign (Ours)	914M	105.1 ms/sample	9.51 samples/s

Table 4: Comparison of inference efficiency and parameter counts on the ACE05 test set.

tary roles in refining the semantic space. Appendix C visualizes its overall impact on co-occurring entity consistency by comparing the constraint loss on ACE05 (*w/* vs. *w/o*).

Inference-Related Strategies. 1) w/o Decoder Aggregation: Removing the bidirectional merge mechanism causes a 1.66% decline. This indicates that the recall gain from recovering true positives outweighs the precision loss incurred by potential false positives, leading to an overall performance improvement. **2) w/o Generation Constraint:** The removal of the generation constraint results in a 1.72% drop. This confirms that the benefits of pruning false positives significantly outweigh the risk of erroneously filtering valid predictions. **3) w/o Aggregation & Constraint:** The simultaneous removal of both aggregation and filtering leads to the most severe degradation of 2.72%. Since these two strategies optimize performance from complementary directions (enhancing recall vs. precision) without conflict, their combined application achieves optimal results.

4.6 Analysis of Complex Structures

To provide a clearer picture of our model's performance on complex structures, we conduct a targeted analysis on the SciERC dataset, which inherently contains both nested entities and overlapping relations. The results on specific test set subsets are summarized in Table 3.

Nested Entities. Although nested structures present a substantial challenge for all models, our method maintains stable performance and slightly outperforms the baseline (+0.16% in RE Strict F1). This demonstrates that our approach helps the model maintain logical consistency even in dense environments with overlapping entity boundaries.

Overlapping Relations. Our method achieves a significant improvement in relation extraction (+2.64% in RE Strict F1) compared to the baseline. This confirms that the anchor-based alignment effectively leverages shared entities to bridge multiple relations, significantly mitigating the bottleneck typically caused by overlapping relations in joint extraction tasks.

4.7 Inference Efficiency

To provide a comprehensive evaluation of our framework, we report the parameter counts, inference latency, and throughput on the ACE05 test set (using a single GPU with a batch size of 40), as summarized in Table 4.

Although our framework deploys three decoders, resulting in an increased overall parameter count (914M), these decoders operate in parallel. Consequently, the inference time does not scale linearly with the number of decoders. Compared to representative single-decoder or discriminative baselines like BiSPN and PL-Marker, our approach inevitably exhibits higher inference latency. We attribute this primarily to the inherently larger parameter scale of the generative backbone (e.g., BART-Large), which is a recognized trade-off in generative IE paradigms aimed at achieving superior extraction performance, rather than a computational bottleneck of our multi-decoder design.

4.8 Case Study

Figure 3 illustrates two representative cases from the Text2DT and SciERC datasets.

Input	Ours		Baseline	
	Entity	Relation	Entity	Relation
Valid case 原发性纤毛运动障碍综合征患者@急性肺部感染时，需及时使用有效抗生素抗感染治疗；对于肺部反复感染者，可以使用化痰药物（如N-乙酰半胱氨酸、福多司坦、羧甲司坦）治疗	**True Pred:** (急性肺部感染, ‹症状›) (抗生素, ‹药物›) (福多司坦, ‹药物›) (N-乙酰半胱氨酸, ‹药物›) (化痰药物, ‹药物›) (羧甲司坦, ‹药物›) (原发性纤毛运动障碍综合征患者, ‹病患›) (肺部反复感染, ‹症状›) **Wrong Pred:** (抗感染治疗, ‹治疗›) **Missing Relation:** ()	**True Pred:** (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 抗生素) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 福多司坦) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, N-乙酰半胱氨酸) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 化痰药物) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 羧甲司坦) (原发性纤毛运动障碍综合征患者, ‹临床表现›, 肺部反复感染) (原发性纤毛运动障碍综合征患者, ‹临床表现›, 急性肺部感染) **Wrong Pred:** () **Missing Relation:** ()	**True Pred:** (急性肺部感染, ‹症状›) (抗生素, ‹药物›) (福多司坦, ‹药物›) (N-乙酰半胱氨酸, ‹药物›) (化痰药物, ‹药物›) (羧甲司坦, ‹药物›) (原发性纤毛运动障碍综合征患者, ‹病患›) (肺部反复感染, ‹症状›) **Wrong Pred:** (抗生素抗感染治疗, ‹治疗›) **Missing Relation:** ()	**True Pred:** (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 福多司坦) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, N-乙酰半胱氨酸) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 化痰药物) (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 羧甲司坦) (原发性纤毛运动障碍综合征患者, ‹临床表现›, 肺部反复感染) (原发性纤毛运动障碍综合征患者, ‹临床表现›, 急性肺部感染) **Wrong Pred:** (原发性纤毛运动障碍综合征患者, ‹治疗方案›, 抗生素抗感染治疗) **Missing Relation:** (原发性纤毛运动障碍综合征患者, ‹治疗药物›, 抗生素)
Failure case It has also been studied in the framework of Japanese information extraction -LRB- -LSB- 3 -RSB- -RRB- in recent years .	**True Pred:** (It, ‹generic›) (Japanese information extraction, ‹task›) **Wrong Pred:** () **Missing Relation:** ()	**True Pred:** () **Wrong Pred:** (It, ‹used-for›, Japanese information extraction) (Japanese information extraction, ‹part-of›, It) **Missing Relation:** (Japanese information extraction, ‹used-for›, It)	**True Pred:** (It, ‹generic›) (Japanese information extraction, ‹task›) **Wrong Pred:** () **Missing Relation:** ()	**True Pred:** () **Wrong Pred:** (It, ‹used-for›, Japanese) (Japanese, ‹part-of›, It) **Missing Relation:** (Japanese information extraction, ‹used-for›, It)

Figure 3: Cases from Text2DT and SciERC.

Valid Case: The baseline suffers from a span nesting issue, predicting an erroneous entity that encapsulates the correct one. This ambiguity triggers a cascading error, leading to the inevitable misallocation of the corresponding tail entity. Although the model trained with semantic alignment does not fully rectify the entity prediction error, it enforces clearer boundary distinctions between overlapping entities. Consequently, the tail entity is correctly identified during the relation extraction phase, demonstrating the efficacy of our method.

Failure Case: Although the proposed semantic alignment successfully rectifies the span boundary for "Japanese information extraction", it fails to recover the relation triplet due to type inconsistency and head-tail reversal. This suggests that structural improvements alone cannot compensate for the lack of intrinsic domain knowledge needed for complex reasoning.

5 Conclusion

In this work, we systematically explore structured output paradigms and introduce AnchorAlign to address the critical misalignment in joint extraction. Specifically, we propose the concept of the anchor entity to unify existing output paradigms, which theoretically enables the derivation of two novel output paradigms. Building on this foundation, we design a semantic alignment mechanism that leverages anchor entities to rigorously model intra-task RE dependencies while effectively mitigating biases arising from inter-task (NER-RE) dependencies. Extensive empirical results across five benchmark datasets confirm that our approach yields significant improvements over competitive baselines, achieving new SOTA performance.

Limitations

While the semantic alignment objective effectively sharpens entity boundaries and facilitates the correct assignment of relation arguments, it offers limited gains in rectifying the semantic correctness of predicted entities, as evidenced by persistent entity encapsulation errors. Furthermore, the model's performance remains constrained by its inherent knowledge base. As demonstrated by the Text2DT case, accurate entity extraction does not automatically translate into accurate relation extraction. The model continues to exhibit misalignments in relation classification and directionality judgment (e.g., head-tail reversal), indicating that alignment constraints cannot fully compensate for deficiencies in the model's reasoning knowledge. In future work, we will attempt to incorporate more dependency information among IE subtasks and explore new avenues for injecting more comprehensive domain knowledge to bolster the model's reasoning capabilities.

Acknowledgments

This study is partially supported by National Natural Science Foundation of China (62276082), National Key RD Program of China (2023YFC3502900), Shenzhen Science and Technology Research and Development Fund (KJZD20240903102802003), Shenzhen Science and Technology Research and Development Fund for Sustainable Development Project (GXWD20231128103819001, 20230706140548006) and Guangdong Provincial Key Laboratory Grant (2023B1212060076).

References

Masaki Asada and Makoto Miwa. 2025. Improving relation extraction by sequence-to-sequence-based dependency parsing pre-training. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7099–7105, Abu Dhabi, UAE. Association for Computational Linguistics.

Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. Rebel: Relation extraction by end-to-end language generation. In *Findings of the association for computational linguistics: emnlp 2021*, pages 2370–2381.

Harsha Gurulingappa, Abdul Mateen Rajput, Angus Roberts, Juliane Fluck, Martin Hofmann-Apitius, and Luca Toldo. 2012. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of Biomedical Informatics*, 45(5):885–892. Text Mining and Natural Language Processing in Pharmacogenomics.

Yuxin He and Buzhou Tang. 2023. Bispn: Generating entity set and relation set coherently in one pass. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2066–2077.

Martin Josifoski, Nicola De Cao, Maxime Peyrard, Fabio Petroni, and Robert West. 2022. GenIE: Generative information extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4626–4643, Seattle, United States. Association for Computational Linguistics.

Hyeonseok Kang, Hyein Seo, Jeesu Jung, Sangkeun Jung, Du-Seong Chang, and Riwoo Chung. 2024. Guidance-based prompt data augmentation in specialized domains for named entity recognition. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 665–672, Bangkok, Thailand. Association for Computational Linguistics.

Peng Li, Tianxiang Sun, Qiong Tang, Hang Yan, Yuanbin Wu, Xuanjing Huang, and Xipeng Qiu. 2023. CodeIE: Large code generation models are better few-shot information extractors. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15339–15353, Toronto, Canada. Association for Computational Linguistics.

Zixuan Li, Yutao Zeng, Yuxin Zuo, Weicheng Ren, Wenxuan Liu, Miao Su, Yucan Guo, Yantao Liu, Xiang Li, Zhilei Hu, and 1 others. 2024. Knowcoder: Coding structured knowledge into llms for universal information extraction. *arXiv preprint arXiv:2403.07969*.

Jie Lou, Yaojie Lu, Dai Dai, Wei Jia, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2023. Universal information extraction as unified semantic matching. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 37, pages 13318–13326.

Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2022. Unified structure generation for universal information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5755–5772, Dublin, Ireland. Association for Computational Linguistics.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.

Giovanni Paolini, Ben Athiwaratkun, Jason Krone, Jie Ma, Alessandro Achille, Rishita Anubhai, Cicero Nogueira dos Santos, Bing Xiang, and Stefano Soatto. 2021. Structured prediction as translation between augmented natural languages. *arXiv preprint arXiv:2101.05779*.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. *Modeling Relations and Their Mentions without Labeled Text*, page 148–163.

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, and Eneko Agirre. 2023. Gollie: Annotation guidelines improve zero-shot information-extraction. *arXiv preprint arXiv:2310.03668*.

C. Walker and Linguistic Data Consortium. 2005. *ACE 2005 Multilingual Training Corpus*. LDC corpora. Linguistic Data Consortium.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2022. DeepStruct: Pretraining of language models for structure prediction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 803–823, Dublin, Ireland. Association for Computational Linguistics.

Xiao Wang, Weikang Zhou, Can Zu, Han Xia, Tianze Chen, Yuansen Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, and 1 others. 2023. Instructuie: Multi-task instruction tuning for unified information extraction. *arXiv preprint arXiv:2304.08085*.

Xinglin Xiao, Yijie Wang, Nan Xu, Yuqi Wang, Hanxuan Yang, Minzheng Wang, Yin Luo, Lei Wang, Wenji Mao, and Daniel Zeng. 2023. Yayi-uie: A chat-enhanced instruction tuning framework for universal information extraction. *arXiv preprint arXiv:2312.15548*.

Hang Yan, Tao Gui, Junqi Dai, Qipeng Guo, Zheng Zhang, and Xipeng Qiu. 2021. A unified generative framework for various NER subtasks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5808–5822, Online. Association for Computational Linguistics.

Deming Ye, Yankai Lin, Peng Li, and Maosong Sun. 2022. Packed levitated marker for entity and relation extraction. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 4904–4917.

Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Xu Ming, and Yefeng Zheng. 2021. PRGC: Potential relation and global correspondence based joint relational triple extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6225–6235, Online. Association for Computational Linguistics.

Wei Zhu, Wenfeng Li, Xiaoling Wang, Wendi Ji, Yuanbin Wu, Jin Chen, Liang Chen, and Buzhou Tang. 2023. *Extracting Decision Trees from Medical Texts: An Overview of the Text2DT Track in CHIP2022*, page 89–102.

A Data Augmentation

The data augmentation process is decomposed into two distinct phases: entity mention rewriting and context rewriting, as illustrated in Figures 7 and 8. The expanded corpus is subsequently synthesized by cross-matching the generated mention variants with the context variants. A critical component of this process is our index-based consistency constraint: we assign a unique index to each entity and strictly enforce that both the entity type associated with a specific index and the existing relations between indexed entities remain invariant throughout the rewriting process. This mechanism ensures the automated generation of high-quality labeled data tailored for joint entity and relation extraction.

Finally, we performed data augmentation on each dataset, capped at a maximum of 10,000 samples. The augmented data was processed via an automated annotation script, retaining only valid instances to construct the final pre-training corpus. Table 5 illustrates the detailed statistics of the original versus augmented data counts.

Dataset	|labels|	Samples			Augment train
		Train	Val	Test	
ACE05	5	10051	2420	2050	9845
ADE(10-fold avg.)	1	3845	427	-	9782
NYT	24	56196	5000	5000	9767
SciERC	7	1861	275	551	9658
Text2DT	7	400	100	-	3144

Table 5: Statistics of Relation Extraction (RE) Datasets

B Output Paradigm Design Comparison

According to the comparative results in Table 6, Paradigm 1 exhibits the lowest average ranking. Conversely, Paradigm 3 outperforms other competitive variants (Paradigms 2 and 4) by offering a more robust balance of stability and accuracy. Driven by these observations, we prioritize Paradigm 3 as our default paradigm.

Viewing Paradigm 3 as an optimized variant of Paradigm 1, we further investigate the correlation between its performance gains and the density of co-occurring entities in the datasets (detailed in Table 7). We measure the density of shared entities using both per-sample and per-entity metrics, which yield consistent dataset rankings.

A distinct positive correlation emerges: datasets with higher shared density correspond to greater F1 improvements for Paradigm 3 over Paradigm 1. This finding validates our design rationale, confirming that the entity-grouping strategy effectively reduces output redundancy, particularly in scenarios with high shared density.

Output Paradigm	ACE05	SciERC	ADE
Paradigm1	66.95	47.22	84.86
Paradigm2	**68.76**	47.24	85.10
Paradigm3	67.05	**47.62**	**85.24**
Paradigm4	67.21	47.01	85.13

Table 6: Relation Boundary F1 performance comparison of different output paradigms on three public datasets.

Metric	ACE05	SciERC	ADE
Sample-level Shared Density			
Shared Head Ratio	4.30%	21.59%	13.76%
Shared Tail Ratio	2.30%	23.74%	21.79%
Average Ratio	**3.30%**	**22.67%**	**17.78%**
Entity-level Shared Density			
Shared Head Ratio	20.89%	35.41%	27.65%
Shared Tail Ratio	11.11%	37.37%	42.02%
Average Ratio	**16.00%**	**36.39%**	**34.84%**
$\triangle$ F1 (vs. Paradigm 1)	**+0.10**	**+0.40**	**+0.38**

Table 7: Impact of Shared Entity Density (Sample- and Entity-level) on the Performance Gain of Paradigm 3 over Paradigm 1.

C Visual Validation of Semantic Alignment Effectiveness

To visually demonstrate the enhancement in anchor consistency brought by cross-task semantic alignment training, we compare the consistency loss trajectories of models trained with (*w/*) and without (*w/o*) this objective. This comparative analysis is conducted across the training, validation, and test splits of the ACE05 dataset, providing concrete evidence that substantiates the efficacy of our proposed method.

We logged the step-wise loss values for both the baseline and the semantic alignment enhanced models throughout the training process. During post-processing, invalid entries—specifically samples devoid of relation annotations which yield zero values—were filtered out. To mitigate high-frequency stochastic fluctuations and highlight underlying optimization trends, we applied a Simple Moving Average (SMA) smoothing technique with a window size of $w = 7$.

As shown in Figure 4, since both models have achieved near-complete convergence on the seen data (training set), the disparity in consistency loss is negligible. However, a comparison of Figures 5 and 6 reveals that the model trained with semantic alignment exhibits significantly lower consistency loss on unseen samples (validation and test sets). This demonstrates that the semantic alignment objective promotes greater consistency and robustness in the outputs of the two tasks, effectively validating the generalization capability of our approach.

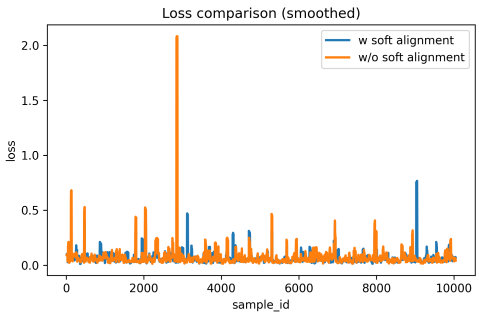

Figure 4: Semantic Alignment Loss (Train)

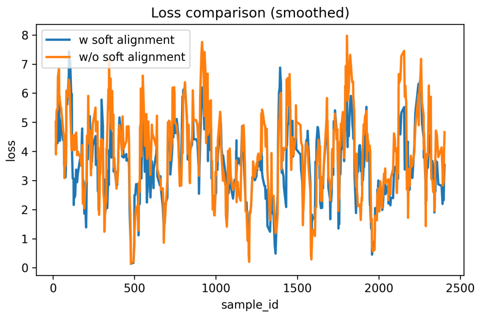

Figure 5: Semantic Alignment Loss (Dev)

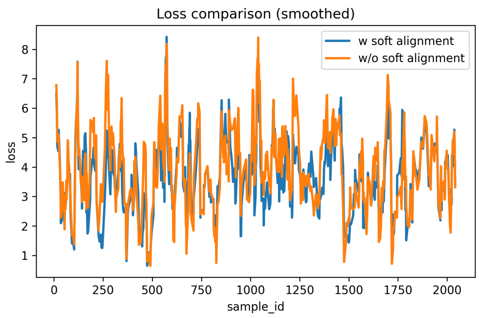

Figure 6: Semantic Alignment Loss (Test)

D Hyperparameter Configuration

For the pre-training stage, we report the exact epochs spent in **Entity Type Understanding** and

Dataset	**Epochs**		**Learning Rate** $(\times 10^{-5})$		**Batch Size**	
	Pre-training	**SFT**	**Pre-training**	**SFT**	**Pre-training**	**SFT**
ACE05	[4, 6, 10]	[10, 35, 55]	3	[2, 2, 2]	25	25
SciERC	[2, 3, 10]	[30, 70, 100]	3	[3, 3, 3]	25	25
ADE	[2, 3, 5]	[10, 35, 55]	2	[2, 2, 2]	30	30
NYT	[2, 3, 15]	[10, 35, 55]	3	[3, 3, 2]	25	25
Text2DT	[2, 3, 10]	[10, 70, 100]	3	[3, 3, 2]	10	15

Table 8: Detailed hyper-parameter settings across Pre-training and Supervised Fine-tuning (SFT).

Progressive Objective Activation (both without alignment and within alignment training) while keeping the learning rate fixed.

For the SFT stage, we list the epochs and corresponding learning rates used in **Curriculum Warm-up** and **Full Data Training** (again, both without alignment and within alignment training). The final hyper-parameter choices for every dataset are provided in Table 8.

Entity Mention Rewriting

Your task is to rewrite the entity mentions represented in dictionary format {"entity_type": entity_mention} within the input sentence, ensuring that the total number of entities remains unchanged while preserving each entity's original order and category. Only the entity mentions themselves should be rewritten—the rewritten mentions must align with the contextual environment and maintain consistent relationships with other entity mentions, while keeping all other parts of the sentence structure intact. Output each version on a separate line.

Input:
Meanwhile , {"peop": "Shi Liming"} at the {"org": "Institute of Zoology"} of {"loc": "Kunming"} found that pandas lack variety in their protein heredity , which may serve as one of the major reasons for pandas ' near extinction .

Output:
1. Meanwhile, {"peop": "a leading researcher"} at the {"org": "Zoological Research Center"} of {"loc": "Yunnan"} found that pandas lack variety in their protein heredity, which may serve as one of the major reasons for pandas' near extinction.
2. Meanwhile, {"peop": "Dr. Shi"} at the {"org": "Chinese Academy of Sciences"} of {"loc": "Kunming City"} found that pandas lack variety in their protein heredity, which may serve as one of the major reasons for pandas' near extinction.
3. Meanwhile, {"peop": "a scientist"} at the {"org": "Wildlife Conservation Institute"} of {"loc": "southwestern China"} found that pandas lack variety in their protein heredity, which may serve as one of the major reasons for pandas' near extinction.

Input:
"It is my sincere hope that we will reignite our united purpose , " Senator {"person": "Mitch McConnell"} of {"location": "Kentucky"} , the Republican whip , said on the Senate floor before the ceremony .

1. "It is my sincere hope that we will reignite our united purpose," Senator {"person": "the GOP leader"} of {"location": "the Bluegrass State"}, the Republican whip, said on the Senate floor before the ceremony.
2. "It is my sincere hope that we will reignite our united purpose," Senator {"person": "McConnell"} of {"location": "KY"}, the Republican whip, said on the Senate floor before the ceremony.
3. "It is my sincere hope that we will reignite our united purpose," Senator {"person": "the Senate Minority Leader"} of {"location": "his home state"}, the Republican whip, said on the Senate floor before the ceremony.

Figure 7: Entity mention rewriting

Context Rewriting

Your task is to rewrite the given sentence into three distinct versions, ensuring that each unique <mask_id> represents a masked entity. The rewritten versions must retain the same number of <mask_id> placeholders, with the contextual environment remaining consistent with the original <mask_id> while preserving the implied relationships between them. Output each version on a separate line.

Input:
Meanwhile , <mask0> at the <mask1> of <mask2> found that pandas lack variety in their protein heredity , which may serve as one of the major reasons for pandas ' near extinction .

Output:
1. Meanwhile, <mask0> at the <mask1> of <mask2> discovered that pandas have limited protein heredity diversity, potentially contributing to their near-extinction status.
2. At the <mask1> of <mask2>, <mask0> found that pandas exhibit a lack of protein heredity variation, which could be a key factor in their near extinction.
3. Pandas' near extinction may be partly due to their low protein heredity variety, as revealed by <mask0> at the <mask1> of <mask2>.

Input:
"It is my sincere hope that we will reignite our united purpose , " Senator <mask0> of <mask1> , the Republican whip , said on the Senate floor before the ceremony .

1. "It is my sincere hope that we will reignite our united purpose," declared Senator <mask0> of <mask1>, the Republican whip, during a speech on the Senate floor prior to the ceremony.
2. Before the ceremony, Senator <mask0> of <mask1>, the Republican whip, expressed on the Senate floor, "It is my sincere hope that we will reignite our united purpose."
3. Speaking on the Senate floor ahead of the ceremony, Senator <mask0> of <mask1>, the Republican whip, stated, "It is my sincere hope that we will reignite our united purpose."

Figure 8: Context rewriting