R1-RE: Cross-Domain Relation Extraction with RLVR

Runpeng Dai[1] Tong Zheng[2] Run Yang[3] Kaixian Yu[4] Hongtu Zhu[1†]

[1]University of North Carolina at Chapel Hill [2]University of Maryland, College Park
[3]University of Alberta [4]Insilicom LLC
{runpeng, htzhu}@email.unc.edu
tzheng24@umd.edu
run6@ualberta.ca
kaixian@insilicom.com

Abstract

Relation extraction (RE) is a core task in natural language processing. Traditional approaches typically frame RE as a supervised learning problem, directly mapping context to labels—an approach that often suffers from poor out-of-domain (OOD) generalization. Inspired by the workflow of human annotators, we reframe RE as a reasoning task guided by annotation guidelines and introduce **R1-RE**, the first reinforcement learning with verifiable reward (RLVR) framework for RE tasks. Our method elicits the reasoning abilities of small language models for annotation tasks, resulting in significantly improved OOD robustness. We evaluate our approach on the public Sem-2010 dataset and a private MDKG dataset. The R1-RE-7B model attains an average OOD accuracy of approximately 70%, on par with leading proprietary models such as GPT-4o. Additionally, our comprehensive analysis provides novel insights into the training dynamics and emergent reasoning behaviors of the RLVR paradigm for RE. [1]

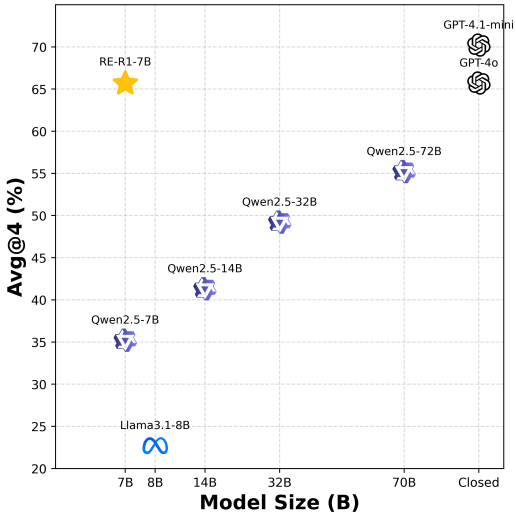


Figure 1: Testing accuracy on the MDKG dataset for R1-RE-7B trained on the Sem-2010 dataset, compared with other models. Detailed results are provided in Table 3.

1 Introduction

Relation extraction (RE) (Zhao et al., 2024) is a fundamental task in natural language processing (NLP) that involves either classifying the relationships between pairs of entities (relationship classification) or extracting (subject, relation, object) triples (triplet extraction) from context. RE serves as a foundation for numerous downstream applications (Nayak et al., 2021), most notably in the construction of knowledge graphs (KGs) (Zhong et al., 2023), which have demonstrated broad utility across diverse domains such as biomedical research (Yang et al., 2025; Liu et al., 2026) and e-commerce (Li et al., 2020). Accordingly, improving the reliability and accuracy of RE techniques is crucial.

Standard approaches to RE typically adopt a pretraining-and-finetuning paradigm, leveraging mid-sized pre-trained models such as BART (Lewis et al., 2019) and BERT (Devlin et al., 2019). More recently, supervised fine-tuning (SFT) of open-source LLMs has been explored for RE tasks (Ettaleb et al., 2025). However, as shown in Figure 2(b), naive SFT improves in-domain performance but offers only limited improvement out of domain. This is because such training approaches tend to focus on memorizing the mapping between sentences and gold standards (Chu et al., 2025) rather than developing genuine annotation ability. Other studies have sought to leverage the in-context learning capabilities of LLMs (Brown et al., 2020) by exploring few-shot prompting strategies (Wadhwa et al., 2023; Xu et al., 2023). However, the effectiveness of in-context learning largely depends on access to strong API models[2]. For smaller models, as illustrated in Figure 2(a), few-shot learning yields only marginal improvements.

[1]https://github.com/RunpengDai/R1_RE

[2]API models refer to proprietary, closed-source models that are typically accessible only via external APIs.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 34387–34401

Figure 2: Performance of Qwen2.5-7B-Instruct. (a) Few-shot performance on two RE datasets. (b) The impact of prompt design on the fine-tuned model's performance, comparing a naive prompt (Figure 3) against an annotation-guided prompt (Figure 4) for both in-domain and out-of-domain test accuracy after SFT.

This observation raises a critical question: *How to enhance the RE capability of small LLMs for robust cross-domain performance?* To this end, we drew inspiration from the workflow of human annotators (see Figure 3). Annotators are typically provided with detailed annotation guidelines and iteratively compare the target sentence against these guidelines—formulating hypotheses, verifying them, and ultimately reaching a conclusion. While entities and relation types may vary across domains, the reasoning skills developed through annotation guideline usage are broadly generalizable.

Human intelligence motivates us to conceptualize RE as a reasoning process anchored in annotation guidelines. However, eliciting such reasoning skills from small language models is challenging. As a first step, we adapt the standard SFT paradigm by explicitly incorporating the annotation guide into the prompt. As shown in Figure 2, this approach yields roughly a 10% boost in out-of-domain RE performance; nevertheless, a substantial gap remains relative to proprietary API models. To address this, we introduce **R1-RE**, a novel framework designed to further enhance the reasoning capabilities of LLMs for RE tasks. Our approach is inspired by recent advances in reinforcement learning with verifiable reward (RLVR), which has demonstrated strong potential for promoting reasoning in smaller models on complex domains such as mathematics and code generation (Guo et al., 2025).

We evaluate our proposed method on both a public dataset (SemEval-2010 Task 8) and Mental Disorder Knowledge Graph (MDKG) dataset (Gao et al., 2025). Our results show that R1-RE improves the OOD performance of Qwen2.5-7B-Instruct by up to **+30 pp**. Notably, our 7B model achieves performance comparable to GPT-4o on MDKG. Further analysis shows that: (1) R1-RE elicits genuinely human-like annotation behavior. (2) Incorporating additional training data can lead to even greater performance gains. (3) The training process of R1-RE preserves the model's performance on other tasks.

2 Preliminaries

2.1 Task Definition

Let $\mathcal{S}$ denote the input sentence, and let $\mathcal{E}$ and $\mathcal{R}$ represent the predefined sets of entity types and relation types, respectively. Let $\mathcal{K}$ denote the annotation guideline for the task, which provides the definitions of $\mathcal{E}$ and $\mathcal{R}$ as well as other annotation instructions. We consider two relation extraction (RE) tasks: Relation Classification (RC) and Triplet Extraction (TE). In this work, we primarily focus on **RC** tasks; leaving the discussions on **TE** task in Appendix D.

Relation Classification (RC): In this task, the subject-object pair (e_{sub}, e_{obj}) has already been identified from $\mathcal{S}$. The goal is to assign a relation type $y \in \mathcal{Y}$, where $\mathcal{Y}$ is the predefined set of relation types, based on the sentence $\mathcal{S}$ and the annotation guideline $\mathcal{K}$. Each sentence $\mathcal{S}$ contains exactly one entity pair. For instance,

Sentence <e1>Counseling interventions</e1> can be effective in preventing <e2>perinatal depression</e2>.
Gold standard treatment-for(e1, e2)

2.2 Group Relative Policy Optimization (GRPO)

In this work, we formulate the language generation process of LLMs as a sequential decision process and optimize the LLM with the Group Relative Policy Optimization (GRPO) algorithm (Shao et al., 2024).

Specifically, let π_θ denote the LLM with parameters θ. At each training step, given a prompt q sampled from the dataset $\mathcal{D}$, we use π_θ to generate a group of G candidate outputs, denoted as $o_1, o_2, \ldots, o_G$. For each candidate output, we compute the corresponding reward $r_1, r_2, \ldots, r_G$ by comparing the output with the gold standard. The advantage at the t-th token of the i-th output is then

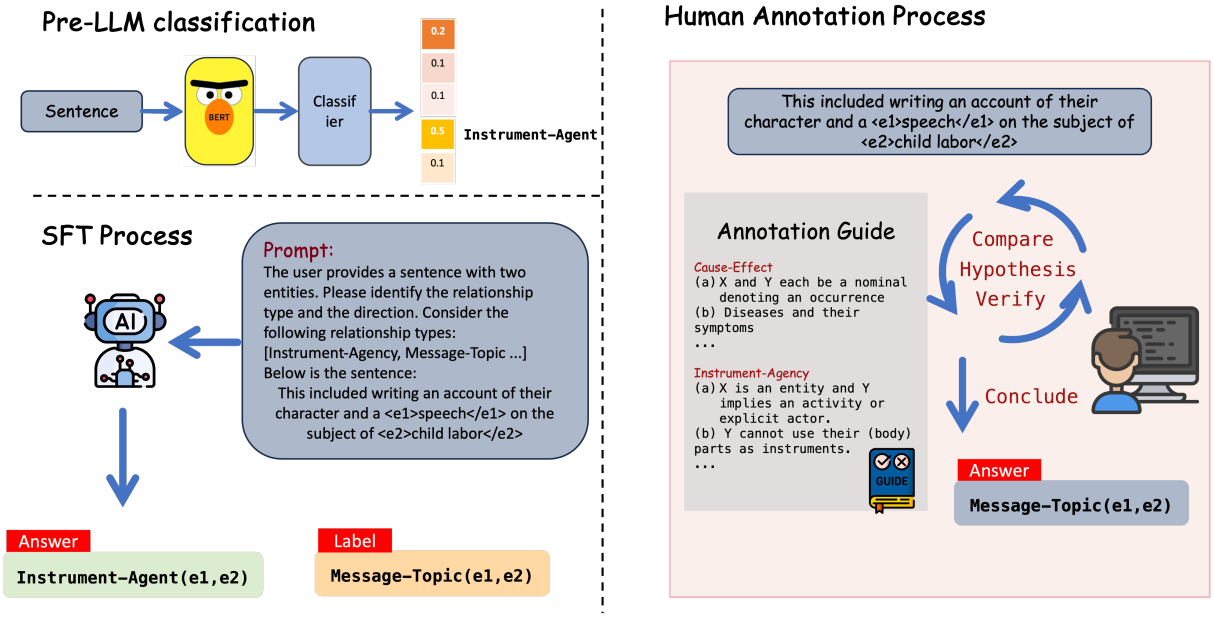


Figure 3: A comparison of existing RE training paradigm and the annotation process of human annotators.

calculated as

$$A_{i,t} = A_i = \frac{r_i - \text{mean}(r_1, r_2, \ldots, r_G)}{\text{std}(r_1, r_2, \ldots, r_G)}.$$

Let $\pi_{\theta_{\text{old}}}$ denote the model from the previous training step, and π_{ref} denote the original model prior to training. GRPO maximizes the following objective function to optimize π_θ:

$$\mathbb{E}_{q\sim\mathcal{D},\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|o_i|}\sum_{t=1}^{|o_i|}\min\Big(\frac{\pi_\theta^{i,t}}{\pi_{\theta_{\text{old}}}^{i,t}}A_i,\ \text{clip}\big(\frac{\pi_\theta^{i,t}}{\pi_{\theta_{\text{old}}}^{i,t}},1-\varepsilon,\ 1+\varepsilon\big)A_i\Big) - \beta\, D_{\text{KL}}\big(\pi_\theta \,\|\, \pi_{\text{ref}}\big)\Bigg],$$

where $\pi_\theta^{i,t} = \pi_\theta(o_{i,t} \mid q, o_{i,1} \ldots, o_{i,t-1})$, and similarly for $\pi_{\theta_{\text{old}}}^{i,t}$. The hyperparameters ε and β control the ratio clipping threshold and the weight of the Kullback–Leibler (KL) divergence penalty, respectively. Specifically, ratio clipping mitigates the risk of large, destabilizing policy updates, while the KL penalty constrains the updated policy from deviating excessively from the reference model π_{ref}. For additional details on the formulation of such sequential decision process and the GRPO algorithm, please refer to (Yuan et al., 2025; Yu et al., 2025).

3 Method

In this section, we first introduce a new paradigm for relation extraction inspired by the human annotation process (Section 3.1). We then present the R1-RE framework (Section 3.2) along with its associated reward design (Section 3.3).

3.1 Human-Inspired RE paradigm

As illustrated in Figure 3, existing relation extraction methods always focus on learning mappings between sentences and gold labels, for example, gold relations or gold standard annotations. Despite achieving strong in-domain performance, these direct mapping strategies often suffer from poor generalization to out-of-domain scenarios.

In contrast, human annotation is inherently a multi-step reasoning process. As shown in Figure 3 (right), annotators consult an annotation guide—a manual that precisely defines each relation and entity type and provides detailed instructions for edge cases (see Table 2). During annotation, they repeatedly compare the target sentence with these definitions, formulate hypotheses, and iteratively refine their judgments through several inference steps before reaching a final decision. While concrete definitions may vary across tasks, this step-by-step reasoning paradigm is universal, making it widely applicable to diverse RE settings. Motivated by this insight, we introduce a RE framework that explicitly embeds the step-by-step human annotation process into LLM-based relation extraction, thereby bridging the gap between human reasoning and automated learning. Specifically, we introduce the following prompt, designed to guide LLMs to-

ward human-style reasoning in relation extraction.

> The user provides a sentence containing two entities: one enclosed in <e1> </e1> tags and the other in <e2> </e2> tags. Please identify the relation type and determine the direction of the relation between these two entities. Below are the definitions and restrictions for all relation types:
> {Annotation guide}
> Please think about the reasoning process in the mind and then provides the user with the final answer. The reasoning process and final class are enclosed within <think> </think> and <answer> </answer> tags, respectively. The answer should align with format: For example, <answer> Product-Producer(e1,e2) </answer> means e1 is a product of e2 while <answer> Product-Producer(e2,e1) </answer> means e2 is a product of e1. Always use "e1" and "e2" in place of the actual entity names. Below is the sentence:
> {Sentence}

Figure 4: The prompt for RC tasks.

3.2 R1-RE

Our goal is to elicit human-like annotation behavior in the LLM-based RE pipeline. Achieving this, however, is far from trivial. Preliminary experiments show that simply prompting an LLM to emulate the annotation procedure yields unsatisfactory results. As illustrated in Figure 5, the model drifts toward a shallow, linear chain of thought instead of the richer, multi-step reasoning observed in human annotators.

To this end, we propose R1-RE, a R1-style reinforcement learning training framework that aligns an LLM's reasoning with the multi-step workflow of human annotators. This is motivated by the recent evidence that reinforcement learning with verifiable reward, as discussed in Section 5, can successfully equip a LLM with the capability of performing human-like reasoning on a wide range of tasks including math (Shao et al., 2024), coding and QA (Lai et al., 2025).

Given an LLM policy π, our goal is to maximize the expected reward:

$$\mathbb{E}_{(q,y)\sim\mathcal{D},\widehat{y}\sim\pi(\cdot|q)}\left[r(\widehat{y},y)\right] \quad (1)$$

where $\mathcal{D}$ is the dataset of prompt–gold label pairs, q denotes a prompt, y the corresponding gold-standard RE result, and $\widehat{y}$ the model-generated output conditioned on q. The reward function $r(\widehat{y},y)$ measures the quality of the generated answer relative to the gold label. We optimize π with respect to this objective using the GRPO algorithm (Shao et al., 2024), as described in Subsection 2.2.

3.3 Multi-stage Reward Design

Reward design is a critical aspect of reinforcement learning, as it encodes feedback from the environment to guide the learning process. In this work, we adopt a rule-based reward scheme similar to those used in DeepSeek-R1-Zero (Guo et al., 2025) and Logic R1 (Xie et al., 2025). Our reward consists of two components. The first is a format reward, which checks whether the model's response adheres to a specified structure. Since the primary purpose of this component is to facilitate evaluation, we enforce only minimal requirements on the response format. The second component is the accuracy reward. For relation classification, we employ a binary reward—assigning positive feedback for correct predictions and negative feedback for incorrect ones. In this subsection, we discuss the reward design for the **RC** task; details for the **TE** task are provided in the Appendix.

Format Reward: The structured prompt template for RC is illustrated in Figure 4. We use regular expression extraction to enforce a standardized response format. The model is permitted to reason freely, without restrictions on the number or placement of `<think></think>` tags. Correctness is determined solely based on the content within the last pair of `<answer></answer>` tags. Furthermore, we require the answer to be in the form $y(e_1, e_2)$ or $y(e_2, e_1)$, where $y \in \mathcal{Y}$ denotes one of the predefined relation types. The format reward score r_{format} is computed as:

$$r_{\text{format}} = \begin{cases} 1, & \text{if format is correct} \\ -3, & \text{if format is incorrect} \end{cases}$$

Metric Reward: If the response passes the format evaluation, we assign an additional reward based on its RE accuracy. Specifically, we employ a rule-based scheme in which the model receives a positive reward for correct classifications and a negative reward for incorrect ones:

$$r_{metric} = \begin{cases} 2, & \text{if } y_{\text{true}} = y_{\text{pred}} \\ -1.5, & \text{otherwise} \end{cases}$$

Final Reward: The final reward r combines a *format reward* (r_{format}) and a *task-specific metric reward* (r_{metric}), and is defined as follows:

$$r = \begin{cases} r_{\text{format}}, & \text{if } r_{\text{format}} \neq 1 \\ r_{\text{format}} + r_{\text{metric}}, & \text{if } r_{\text{format}} = 1 \end{cases}$$

To sum up, our reward design utilizes a three-tiered structure based on the model's response. It assigns a reward of -3 for incorrectly formatted responses, -0.5 for responses with a correct format but a wrong answer, and $+3$ for responses correct in both format and content. We further explore the impact of these specific reward values in the ablation study in Section 4.4.

4 Experiments on Entity Classification

4.1 Dataset

In this paper, we consider two relation classification datasets. The first is the public SemEval-2010 Task 8 dataset (Sem-2010) (Hendrickx et al., 2019), a widely used benchmark for multi-class relation classification. The second is a human-annotated proprietary corpus constructed for the Mental Disease Knowledge Graph (MDKG). Key statistics of both datasets are summarized in Table 1.

	Sem-2010	**MDKG**
Relations	Component-Whole, Instrument-Agency, etc.	Hyponym of, Located in, Risk factor of, Treatment for, etc.
# of classes	17	17
Train/Test	8,353 / 500	10,033 / 500

Table 1: Relation types and train/test splits for the SemEval-2010 and MDKG datasets. The number of classes accounts for relation directionality. The complete table is provided in Appendix C.3.

Each dataset consists of sentence–relation pairs, along with an annotation guide defining each relation type. Table 2 presents an example sentence–relation pair and the corresponding relation definition.

Sentence:	The <e1>hypothalamus</e1> may as a key brain region involved in the <e2>inflammatory related depressive-like behaviors</e2>.
Relation:	hyponym-of (e1, e2)
Definition of hyponym-of:	This relation can indicate a hierarchical link, with X being a subordinate or specific instance of Y . . . "X is hyponym of Y" has following types: (a) Direct Categorization: . . . (b) Appositive Formulation: . . . (c) Indicators of Inclusion or Example: . . . (d) Alternative Naming: . . .

Table 2: An example sentence–gold standard pair from the MDKG dataset, along with the definition of the corresponding relation from the annotation guide, is shown below. Full relation definitions for both datasets are provided in Appendix C.1.

4.2 Main Results

The main results are presented in Table 3. Full implementation details are provided in Appendix A. Here, R1-RE-7B and R1-RE-8B refers to our proposed methods, with all evaluations conducted in a zero-shot setting using the prompt shown in Figure 4. All results are reported in terms of **Avg@4** accuracy, which denotes the average Pass@1 accuracy across four samples. The key observations are as follows:

- The reinforcement learning process significantly enhances the relation extraction capabilities of the base model. R1-RE-7B substantially outperforms its backbone (Qwen-2.5-7B-Instruct) in both in-domain ($+52.9, 53.2$) and out-of-domain ($+27.9, 30.6$) accuracy.
- Compared to its SFT counterpart, R1-RE demonstrates much stronger out-of-domain generalization, surpassing it by $+16.0$ and $+15.2$ points.
- Both proprietary and open-source models achieve substantially higher accuracy on the Sem-2010 (public) dataset than on the MDKG (private) dataset, suggesting that the public benchmark may suffer from data leakage.
- R1-RE models demonstrate OOD performance on the private MDKG dataset that is comparable to state-of-the-art proprietary models such as GPT-4o and GPT-4.1-mini. The relatively lower performance observed on the Sem-2010 dataset is likely attributable to data leakage, which artificially inflates the accuracy of proprietary models.

4.3 How R1-RE Boosts RE Performance?

In this section, we further investigate a key question: *How R1-RE improves RE performance?* To answer this question, we visualize the training dynamics of R1-RE and conduct a human analysis of its outputs. Figures 6 in Appendix B track response length, training reward, and both in-domain and out-of-domain (OOD) accuracy over the course of training, while Figure 5 compares outputs from the Qwen-2.5-7B-Instruct and R1-RE. We observe two key findings:

- **Key Finding 1: Performance is correlated with both training rewards and response length.** As

Model	MDKG (Private)		Sem-2010 (Open)	
	Avg@4	Pass@4	Avg@4	Pass@4
Proprietary Models				
Claude 3.5 Sonnet	71.1 ± 0.1	99.2	81.3 ± 0.1	100
GPT-4o	65.9 ± 0.1	99.4	79.8 ± 0.0	100
GPT-4.1-mini	70.2 ± 0.1	99.6	81.0 ± 0.2	99.8
Open Source Models				
Qwen-2.5-7B-Instruct	35.2 ± 1.2	48.4	38.6 ± 1.2	59.8
Llama-3.1-8B-Instruct	22.7 ± 0.0	53.8	25.6 ± 0.1	56.2
Qwen-2.5-14B-Instruct	41.3 ± 0.8	56.2	54.6 ± 1.4	74.2
Qwen-2.5-32B-Instruct	49.2 ± 2.1	64.8	66.4 ± 1.2	83.8
Qwen-2.5-72B-Instruct	55.2 ± 1.5	73.6	70.2 ± 0.9	85.4
RL from Instruct Models				
Qwen-2.5-7B-Instruct				
↳ R1-RE-7B (MDKG)	88.1 ± 0.3	90.8	**66.5 ± 0.2**	**84.2**
↳ R1-RE-7B (Sem)	**65.8 ± 0.1**	**76.6**	91.8 ± 0.0	100
Llama-3.1-8B-Instruct				
↳ R1-RE-8B (MDKG)	84.3 ± 0.4	88.7	**67.6 ± 0.1**	**85.0**
↳ R1-RE-8B (Sem)	**61.7 ± 0.1**	**74.5**	90.5 ± 0.1	100
SFT from Instruct Models				
Qwen-2.5-7B-Instruct				
↳ SFT-7B (MDKG)	90.4± 0.2	99.0	**51.3± 0.4**	**69.0**
↳ SFT-7B (Sem)	**49.8± 0.1**	**65.3**	92.4 ± 0.0	100
Llama-3.1-8B-Instruct				
↳ SFT-8B (MDKG)	85.4± 0.2	94.5	**44.6± 0.6**	**64.1**
↳ SFT-8B (Sem)	**46.8± 0.3**	**60.0**	93.4 ± 0.0	100

Table 3: Zero-shot relation classification accuracy of different models on the MDKG and Sem-2010 datasets. **R1-RE-7B (MDKG)** denotes the model trained on the MDKG dataset; the same naming convention applies to the SFT models. **Avg@4** indicates the average Pass@1 accuracy across 4 samples. For fair comparison, all the models use the template in 4. Out-of-domain accuracies are highlighted in **bold**.

illustrated in Figures 6 in Appendix B, training rewards, response length, in-domain accuracy, and out-of-domain (OOD) accuracy all increase simultaneously throughout training. Specifically, compared with around 200 tokens of output at the beginning of the training process, the response length increases to 500/1000 tokens on two datasets, indicating the emergence of long COT and aligns well with the phenomena in the existing literature (Xie et al., 2025).

- **Key Finding 2: R1-RE elicits genuinely human-like annotation behavior.** Unlike the baseline's terse, pattern-matching answers, R1-RE-7B first identifies the entities in the context, then systematically compares each candidate relation with the definitions in the annotation guide, following a hypothesis–validation procedure; it finally draws a conclusion and outputs the answer. This step-by-step reasoning process closely mirrors the human annotation workflow depicted in Figure 3(b). This also indicates that the increased length shown in Figure 6 in Appendix B is indeed a result of learning good annotation paradigms instead of simple overthinking (Sui et al., 2025). Notably, this reasoning pattern emerges naturally during the RL training process, without the need for explicit distillation or supervised fine-tuning (Zheng et al., 2025).

4.4 Further Analysis

- **The training process of R1-RE preserves the model's performance on other tasks.** Although R1-RE has acquired strong RC skills through reinforcement learning, its performance on other tasks remains unclear. Specifically, we are interested in whether the RC training process comes at a cost, such as reduced generalization or capability in unrelated domains. As highlighted in

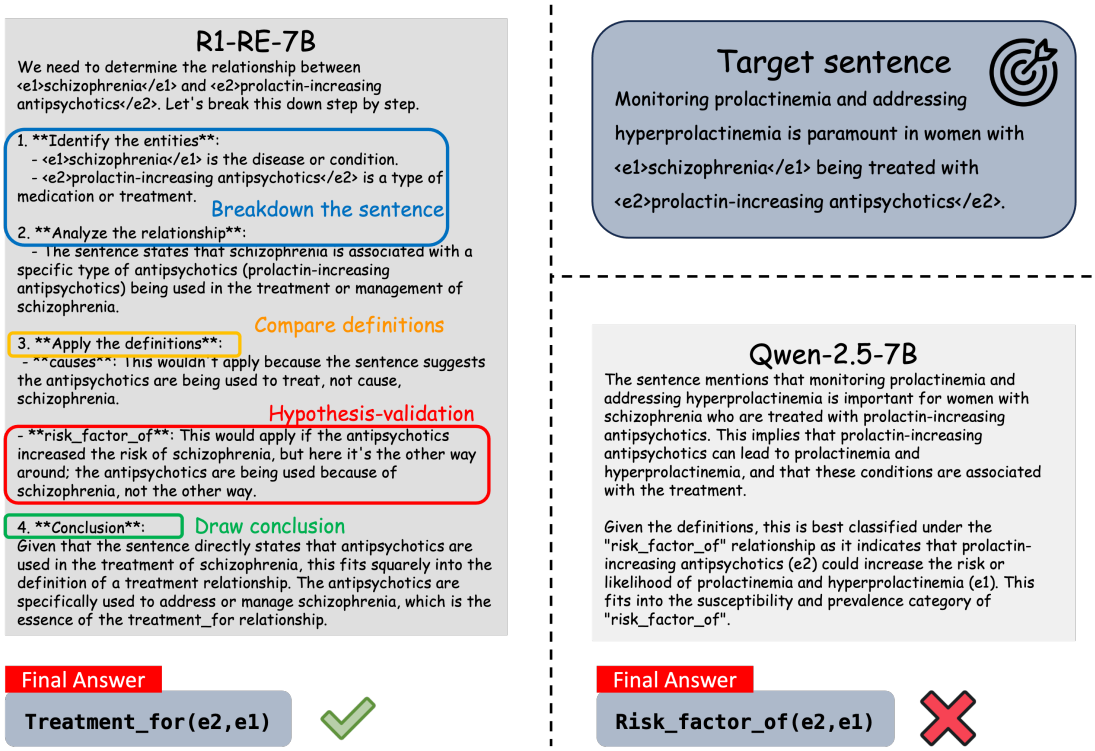


Figure 5: Case study comparing the chain-of-thought (COT) reasoning of **R1-RE-7B** and Qwen2.5-7B-Instruct. Due to space constraints, some COT outputs are omitted; the complete COT reasoning process for **R1-RE-7B** is provided in Appendix C.2.

recent studies (Zhao et al.; Kotha et al., 2023; Dai et al., 2025), when LLMs are fine-tuned to excel at a particular skill, they may struggle to maintain performance in other areas. To investigate this trade-off, we benchmark **R1-RE** on three widely used datasets that evaluate distinct aspects of LLM ability: MATH-500 (mathematical reasoning) (Hendrycks et al., 2021), IFEval (instruction following) (Zhou et al., 2023), and GPQA (factual knowledge) (Rein et al., 2024). The results in Table 4 reveal a surprising yet reasonable trend. The RL process not only failed to suppress but actually improved the performance of R1-RE across all three benchmarks. In contrast, supervised fine-tuning (SFT) led to a noticeable decline in generalization performance. This finding is consistent with recent studies indicating that SFT tends to promote memorization of training data, thereby impairing out-of-domain generalization, whereas RL-based methods can facilitate better skill transfer and generalization across diverse tasks (Chu et al., 2025).

- **Incorporating additional training data can further improve the OOD performance of R1-RE.** Given the strong out-of-domain generalization exhibited by **R1-RE** on RC tasks, a natural question arises: how can we further improve its performance to match or even surpass that of proprietary models?

Model	MATH 500	IFEval	GPQA
Qwen-2.5-7B	73.6	72.3	29.2
R1-RE-7B	75.4	73.0	29.6
SFT	69.8	71.9	27.2

Table 4: Performance of Qwen-2.5-7B-Instruct, **R1-RE**, and SFT models on three benchmarks. The evaluation is conducted with the default setting of the Evalscope framework.

Leveraging the broad availability of public RC datasets, we investigate the effect of incorporating additional data into training. Specifically, we include the SemEval-2018 Task 7 dataset (Buscaldi et al., 2017) (denoted as Sem-2018), training R1-RE on the combined dataset and re-evaluating its out-of-domain (OOD) accuracy. Remarkably, the inclusion of Sem-2018 leads to a substantial 4% improvement in OOD performance on both the MDKG and Sem-2010 datasets. This result demonstrates the significant potential of our approach to further enhance

generalization by integrating diverse and complementary datasets.

Model	Avg@4 (MDKG)	Avg@4 (Sem-2010)
Qwen-2.5-7B-Instruct	35.2	38.6
↳ R1-RE-7B (MDKG)	\	66.5
⊕ Sem-2018	\	**70.8**
↳ R1-RE-7B (Sem-2010)	65.8	\
⊕ Sem-2018	**69.7**	\

Table 5: Out-of-domain (OOD) performance of R1-RE-7B after incorporating the additional Sem-2018 dataset.

- **Robustness to Reward Design.** We conducted an ablation study to analyze the model's sensitivity to the reward function. In this experiment, we tested an alternative reward scheme: a reward of -3 for responses with an incorrect format, -1 for those with the correct format but an incorrect answer, and $+3$ for a correct answer in the correct format. As detailed in Table 6, this modification to the reward scale had a negligible impact on final performance, demonstrating the model's robustness.

Model	Avg@4 (MDKG)	Avg@4 (Sem-2010)
R1-RE-7B (MDKG)	89.2 (+1.1)	67.0 (+0.5)
R1-RE-7B (Sem-2010)	64.7 (-1.1)	90.0 (-1.8)

Table 6: Ablation study on the reward design for **R1-RE-7B** using an alternative reward scale of $\{-3, -1, 3\}$. Performance changes relative to the baseline reward design from Section 3.3 are highlighted in green (improvement) and red (decline).

5 Related works

5.1 LLM Reasoning

LLM reasoning has garnered significant attention, as it demonstrates the potential of LLMs to generalize to complex real-world problems through human-like reasoning in many domains (Tian et al., 2025; Ma et al., 2026). Early efforts primarily focused on prompting methods, such as "chain of thought"(Wei et al., 2022) and "tree of thought"(Yao et al., 2023), to elicit step-by-step reasoning. More recently, research has shifted toward explicitly training LLMs to master reasoning processes. Initial approaches often relied on reward models—such as outcome-based (ORM) or process-based (PRM) reward models (Uesato et al., 2022)—but these methods can suffer from issues like reward hacking.

Recent advances, such as DeepSeek-R1 (Guo et al., 2025) and Vision-R1 (Huang et al., 2025), have demonstrated that applying Reinforcement Learning with Verifiable Reward (RLVR) can effectively guide LLMs toward self-emergent reasoning without requiring trained reward functions or step-level human annotation. RLVR has been explored across a variety of domains, including logic games (Xie et al., 2025), search (Jin et al., 2025), and machine translation (Feng et al., 2025). Also, many efforts in improving (Huang et al., 2026). However, its application to knowledge extraction or relation extraction tasks remains underexplored.

5.2 Relation Extraction

Early approaches to relation extraction predominantly adopt a supervised classification framework. Pipeline-based methods follow a two-step procedure: first applying named entity recognition (NER), then performing relation classification on the identified entity pairs (Cai et al., 2016). Alternatively, span-based methods frame RE as a token-level classification task (Eberts and Ulges, 2020). With the advent of LLMs, recent work has begun to leverage their in-context learning capabilities through few-shot learning (Borchert et al., 2023; Xu et al., 2023). Other approaches seek to enhance performance using retrieval-augmented generation (RAG) methods (Wan et al., 2023). In addition, several studies aim to further improve accuracy via supervised fine-tuning (Wadhwa et al., 2023; Ettaleb et al., 2025; Shi and Luo, 2024). As a fundamental information extraction task, relation extraction plays a crucial role in knowledge graph construction and supports a wide range of biomedical research applications (Liu et al., 2026).

6 Conclusion

In this work, we revisited the relation extraction task by reframing it as a reasoning process grounded in annotation guidelines. We proposed **R1-RE**, a framework that employs RLVR to strengthen the reasoning abilities of LLMs for RE tasks. Experiments on both public and private datasets demonstrate that R1-RE achieves substantial improvements in out-of-domain performance, highlighting promising directions for enhancing the adaptability of LLMs in relation extraction.

Limitations

In this work, we primarily focus on relation classification (RC) tasks and leave the exploration of more complex triplet extraction (TE) tasks for future research. Additionally, our experiments are limited to 7B and 8B-parameter models due to computational constraints. Evaluating R1-RE with larger models represents an important direction for future work.

References

Philipp Borchert, Jochen De Weerdt, Kristof Coussement, Arno De Caigny, and Marie-Francine Moens. 2023. Core: A few-shot company relation classification dataset for robust domain adaptation. *arXiv preprint arXiv:2310.12024*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and 1 others. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Davide Buscaldi, Anne-Kathrin Schumann, Behrang Qasemizadeh, Haïfa Zargayouna, and Thierry Charnois. 2017. Semeval-2018 task 7: Semantic relation extraction and classification in scientific papers. In *International Workshop on Semantic Evaluation (SemEval-2018)*, pages 679–688.

Rui Cai, Xiaodong Zhang, and Houfeng Wang. 2016. Bidirectional recurrent convolutional neural network for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 756–765.

Tianzhe Chu, Yuexiang Zhai, Jihan Yang, Shengbang Tong, Saining Xie, Dale Schuurmans, Quoc V Le, Sergey Levine, and Yi Ma. 2025. Sft memorizes, rl generalizes: A comparative study of foundation model post-training. *arXiv preprint arXiv:2501.17161*.

Runpeng Dai, Run Yang, Fan Zhou, and Hongtu Zhu. 2025. Breach in the shield: Unveiling the vulnerabilities of large language models. *arXiv preprint arXiv:2504.03714*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Markus Eberts and Adrian Ulges. 2020. Span-based joint entity and relation extraction with transformer pre-training. In *ECAI 2020*, pages 2006–2013. IOS Press.

Mohamed Ettaleb, Véronique Moriceau, Mouna Kamel, and Nathalie Aussenac-Gilles. 2025. The contribution of llms to relation extraction in the economic field. In *The Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*.

Zhaopeng Feng, Shaosheng Cao, Jiahan Ren, Jiayuan Su, Ruizhe Chen, Yan Zhang, Zhe Xu, Yao Hu, Jian Wu, and Zuozhu Liu. 2025. Mt-r1-zero: Advancing llm-based machine translation via r1-zero-like reinforcement learning. *arXiv preprint arXiv:2504.10160*.

Shan Gao, Kaixian Yu, Yue Yang, Sheng Yu, Chenglong Shi, Xueqin Wang, Niansheng Tang, and Hongtu Zhu. 2025. Large language model powered knowledge graph construction for mental health exploration. *Nature Communications*, 16(1):7526.

Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, and 1 others. 2024. A survey on llm-as-a-judge. *arXiv preprint arXiv:2411.15594*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid O Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2019. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. *arXiv preprint arXiv:1911.10422*.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*.

Fanding Huang, Guanbo Huang, Xiao Fan, Yi He, Xiao Liang, Xiao Chen, Qinting Jiang, Faisal Nadeem Khan, Jingyan Jiang, and Zhi Wang. 2026. Semantic-space exploration and exploitation in rlvr for llm reasoning. *Preprint*, arXiv:2509.23808.

Wenxuan Huang, Bohan Jia, Zijie Zhai, Shaosheng Cao, Zheyu Ye, Fei Zhao, Zhe Xu, Yao Hu, and Shaohui Lin. 2025. Vision-r1: Incentivizing reasoning capability in multimodal large language models. *arXiv preprint arXiv:2503.06749*.

Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan Arik, Dong Wang, Hamed Zamani, and Jiawei Han. 2025. Search-r1: Training llms to reason and leverage search engines with reinforcement learning. *arXiv preprint arXiv:2503.09516*.

Suhas Kotha, Jacob Mitchell Springer, and Aditi Raghunathan. 2023. Understanding catastrophic forgetting in language models via implicit inference. *arXiv preprint arXiv:2309.10105*.

Yuxiang Lai, Jike Zhong, Ming Li, Shitian Zhao, and Xiaofeng Yang. 2025. Med-r1: Reinforcement learning for generalizable medical reasoning in vision-language models. *arXiv preprint arXiv:2503.13939*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Feng-Lin Li, Hehong Chen, Guohai Xu, Tian Qiu, Feng Ji, Ji Zhang, and Haiqing Chen. 2020. Alimekg: Domain knowledge graph construction and application in e-commerce. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2581–2588.

Dingyuan Liu, Qiannan Shen, and Jiaci Liu. 2026. The health-wealth gradient in labor markets: Integrating health, insurance, and social metrics to predict employment density. *Computation*, 14(1):22.

Weijian Ma, Shizhao Sun, Tianyu Yu, Ruiyu Wang, Tat-Seng Chua, and Jiang Bian. 2026. Thinking with blueprints: Assisting vision-language models in spatial reasoning via structured object representation. *arXiv preprint arXiv:2601.01984*.

Tapas Nayak, Navonil Majumder, Pawan Goyal, and Soujanya Poria. 2021. Deep neural approaches to relation triplets extraction: a comprehensive survey. *Cognitive computation*, 13(5):1215–1232.

Andrew Y Ng, Daishi Harada, and Stuart J Russell. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 278–287.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. 2024. Gpqa: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Y Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Yongliang Shen, Xinyin Ma, Yechun Tang, and Weiming Lu. 2021. A trigger-sense memory flow framework for joint entity and relation extraction. In *Proceedings of the web conference 2021*, pages 1704–1715.

Zhengpeng Shi and Haoran Luo. 2024. Cre-llm: a domain-specific chinese relation extraction framework with fine-tuned large language model. *arXiv preprint arXiv:2404.18085*.

Yang Sui, Yu-Neng Chuang, Guanchu Wang, Jiamu Zhang, Tianyi Zhang, Jiayi Yuan, Hongyi Liu, Andrew Wen, Shaochen Zhong, Hanjie Chen, and 1 others. 2025. Stop overthinking: A survey on efficient reasoning for large language models. *arXiv preprint arXiv:2503.16419*.

Shi-Yu Tian, Zhi Zhou, Wei Dong, Kun-Yang Yu, Ming Yang, Zi-Jian Cheng, Lan-Zhe Guo, and Yu-Feng Li. 2025. Tabularmath: Understanding math reasoning over tables with large language models. *arXiv preprint arXiv:2505.19563*.

Jonathan Uesato, Nate Kushman, Ramana Kumar, Francis Song, Noah Siegel, Lisa Wang, Antonia Creswell, Geoffrey Irving, and Irina Higgins. 2022. Solving math word problems with process-and outcome-based feedback. *arXiv preprint arXiv:2211.14275*.

Somin Wadhwa, Silvio Amir, and Byron C Wallace. 2023. Revisiting relation extraction in the era of large language models. In *Proceedings of the conference. Association for Computational Linguistics. Meeting*, volume 2023, page 15566.

Zhen Wan, Fei Cheng, Zhuoyuan Mao, Qianying Liu, Haiyue Song, Jiwei Li, and Sadao Kurohashi. 2023. Gpt-re: In-context learning for relation extraction using large language models. *arXiv preprint arXiv:2305.02105*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Tian Xie, Zitian Gao, Qingnan Ren, Haoming Luo, Yuqian Hong, Bryan Dai, Joey Zhou, Kai Qiu, Zhirong Wu, and Chong Luo. 2025. Logic-rl: Unleashing llm reasoning with rule-based reinforcement learning. *arXiv preprint arXiv:2502.14768*.

Xin Xu, Yuqi Zhu, Xiaohan Wang, and Ningyu Zhang. 2023. How to unleash the power of large language models for few-shot relation extraction? *arXiv preprint arXiv:2305.01555*.

Yue Yang, Kaixian Yu, Shan Gao, Sheng Yu, Di Xiong, Chuanyang Qin, Huiyuan Chen, Jiarui Tang, Niansheng Tang, and Hongtu Zhu. 2025. Alzheimer's disease knowledge graph enhances knowledge discovery and disease prediction. *Computers in Biology and Medicine*, 192:110285.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, and 1 others. 2025. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*.

Yufeng Yuan, Yu Yue, Ruofei Zhu, Tiantian Fan, and Lin Yan. 2025. What's behind ppo's collapse in long-cot? value optimization holds the secret. *arXiv preprint arXiv:2503.01491*.

Yuexiang Zhai, Christina Baek, Zhengyuan Zhou, Jiantao Jiao, and Yi Ma. 2022. Computational benefits of intermediate rewards for goal-reaching policy learning. *Journal of Artificial Intelligence Research*, 73:847–896.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, and 1 others. A survey of large language models.

Xiaoyan Zhao, Yang Deng, Min Yang, Lingzhi Wang, Rui Zhang, Hong Cheng, Wai Lam, Ying Shen, and Ruifeng Xu. 2024. A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Computing Surveys*, 56(11):1–39.

Tong Zheng, Lichang Chen, Simeng Han, R Thomas McCoy, and Heng Huang. 2025. Learning to reason via mixture-of-thought for logical reasoning. *arXiv preprint arXiv:2505.15817*.

Lingfeng Zhong, Jia Wu, Qian Li, Hao Peng, and Xindong Wu. 2023. A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys*, 56(4):1–62.

Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. 2023. Instruction-following evaluation for large language models. *arXiv preprint arXiv:2311.07911*.

A Training Details

We conduct our experiments on R1-RE models using the `verl` training framework[3], while training the SFT baselines with `LLamMA-Factory` framework [4]. Detailed training configurations for both methods are provided in Table 7. All experiments use full-parameter, full-precision tuning (i.e., no LoRA or quantization). Training was conducted on a cluster with eight NVIDIA A100 (80 GB) GPUs.

B Training Dynamics

Figure 6 provides a comprehensive summary of the training dynamics. The plots track the progression of the training reward, the model's generalization performance on both in-domain and out-of-domain data, and the evolution of the generated response length over time, offering a complete picture of the model's behavior during training.

[3] https://github.com/volcengine/verl

[4] https://github.com/hiyouga/LLaMA-Factory

Parameter	R1-RE	SFT baselines
learning rate	1×10^{-6}	5×10^{-6}
train_batch_size	32	16
total_training_steps	400	600
max_response_length	3K	3K
rollout.n	16	–

Table 7: Hyperparameter comparison between R1-RE and SFT baselines.

C Additional Materials on RC task

C.1 Annotation Guide

The annotation guide for the MDKG dataset is provided in Figure 7, while the guidelines for the SemEval datasets are available on their official website.

C.2 Complete reasoning output from R1-RE-7B and Qwen-2.5-7B-Instruct

The COT reasoning outputs for an example prompt from **R1-RE-7B** and Qwen-2.5-7B-Instruct are shown in Figure 8 and Figure 9, respectively.

C.3 Key statistics of Sem-2010 and MDKG datasets

The complete table of the key statistics of Sem-2010 and MDKG datasets is provided in in Table 8.

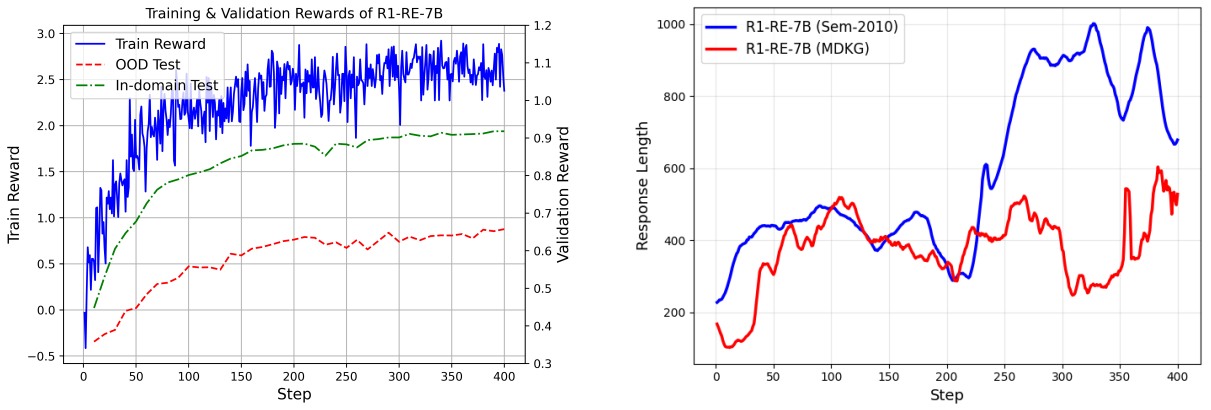
D Additional Materials on TE task

Triplet Extraction(TE): In this task, the goal is to extract all valid triplets (e_{sub}: t_{sub}, y, e_{obj}: t_{obj}) from the sentence $\mathcal{S}$, where $t_{\text{sub}}, t_{\text{obj}} \in \mathcal{E}$ denote the entity types of the subject and object, respectively. A sentence $\mathcal{S}$ may contain multiple triplets. For instance

`Sentence` Olanzapine was also associated with more frequent reports of weight gain and significantly greater VA costs ...
`Label` [[Olanzapine:drug, risk-factor-of, weight gain:symptom]]

D.1 Reward Design

Format Reward: For the TE task, the answer should be a list of triplets, each in the format



(a) (b)

Figure 6: (a) Training dynamics of R1-RE (Sem-2010), with the left y-axis representing the training reward and the right y-axis showing both in-domain and out-of-domain test accuracy. (b) Response length of R1-RE-7B models during training.

	Sem-2010	**MDKG**
Relations	Component-Whole, Cause-Effect Instrument-Agency , Content-Container Message-Topic , Product-Producer Entity-Origin , Member-Collection Other (no direction)	Hyponym of , Located in Characteristic of , Abbreviation for Risk factor of , Occurs in Treatment for , Help diagnose Associated with (no direction)
# of classes	17	17
Train/Test	8,353/ 500	10,033/ 500

Table 8: Relation types and sample sizes of Sem-2010 and MDKG datasets. Number of classes consider the directions of relations.

$(e_{\text{sub}}{:}t_{\text{sub}}, y, e_{\text{obj}}{:}t_{\text{obj}})$, where $y \in \mathcal{Y}$ is a valid relation and $t_{\text{sub}}, t_{\text{obj}} \in \mathcal{E}$ are the corresponding entity types.

Metric Reward: The reward design for the **TE** task is inherently more complex, as it incorporates aspects of the named entity recognition (NER) task, requiring an evaluation of the extracted entities against gold standards.

Prior to the emergence of large language models (LLMs), traditional approaches typically adopted a strict criterion: an entity was considered correct only if the predicted span exactly matched the ground truth (Eberts and Ulges, 2020). However, such evaluation protocols can be overly rigid when applied to LLMs. As noted by Wadhwa et al. (2023), this strictness often fails to account for the more flexible and nuanced outputs of LLMs. To address this limitation, Wadhwa et al. (2023) employed human judges to provide a more accurate and fair assessment of model predictions. In this work, we allow the extracted entity to deviate by one token, either in the front or back, from the gold standard. We propose that a more flexible rule can be adopted, such as using a LLM as the judge (Gu et al., 2024).

We propose a two-stage, F1 score-based as follows:

$$r_{\text{TE}} = w_{\text{ent}} \cdot \text{F1}_{\text{ent}} + w_{\text{tri}} \cdot \text{F1}_{\text{tri}}.$$

Specifically, F1_{ent} denotes the F1 score computed over the extracted entities. To compute F1_{ent}, we extract all unique (entity, entity type) pairs from both the ground truth and the model's output, and calculate the precision and recall. F1_{tri} represents the F1 score over the extracted triplets. For F1_{tri}, we directly compare the predicted and gold-standard triplets.

We make two key observations regarding our reward design. First, the RL framework enables us to directly employ the F_1 score as the reward, thereby eliminating the training–evaluation gap that arises

in conventional approaches which rely on surrogate loss functions (Eberts and Ulges, 2020; Shen et al., 2021). Second, our reward comprises two levels: the entity-level reward acts as an intermediate signal, providing early-stage feedback when the model is not yet capable of extracting complete triplets. This approach is supported both in theory and practice, as intermediate rewards are known to accelerate learning and enhance exploration in RL settings (Ng et al., 1999; Zhai et al., 2022). The triplet-level F_1 score, on the other hand, reflects the final objective of the task. To ensure that learning remains focused on the end goal, we assign higher weight to the triplet-level reward ($w_{\text{ent}} = 1$, $w_{\text{tri}} = 3$), thereby mitigating the risk of divergence from the main objective.

E Prompt Design

The user gives a sentence. The Assistant need to extract triplet from the sentence. Only consider the following entity types:
{Annotation guide - Entity}
Only consider the following relation types:
{Annotation guide - Relation}
The assistant first thinks out load and then provides the user with the final answer, make sure the final answer is enclosed within <answer> </answer> and only appear once. i.e., reasoning process here <answer> answer here </answer>. The answer should be a list of triplets, with each triplet have form [object:type, relation, subject:type].
User: {Sentence}

Figure 10: The prompt for TE tasks.

risk-factor-of: The relation described in S entails that the presence of X heightens the risk or possibility of Y. "X is risk-factor-of Y" has following types:
(a) Direct Risk Factor: The statement "X is a risk factor for Y" indicates that X directly increases the risk of Y occurring.
(b) Cause and Effect: Statements like "X is a leading cause for Y" or "X is caused/induced by Y" emphasize X's causative role in leading to Y. Relevant keywords include "contribute to" and "result in".
(c) Susceptibility and Prevalence: Phrases like "X is prone to developing Y", "X is more prevalent in Y", "X predisposes to Y", "X is at high risk for Y", "higher likelihood of X", "increased odds of X" suggest that X increases the likelihood of Y occurring. Additionally, statements like "X (a specific characteristic group) has an increased Y" fall under this category.
(d) Predictive Relation: The phrase "The happening of X is a predictor of Y" indicates that the occurrence of X can be used to predict the likelihood of Y happening in the future.
(e) Adverse Effects: X is a treatment or intervention and Y is an adverse effect or complication of that treatment.
help-diagnose: This relation type encompasses methods used for diagnosing diseases as well as for observing specific signs, genes, or health factors, and may include diagnostic markers. "X help-diagnose Y" has following types:
(a) Methods of Measurement and Detection: Tag sentences like "Use X to estimate/measure/detect/evaluate/assess Y". Here, X is used as a tool or method to quantify, identify, or assess Y. Other important keywords include 'based on', 'according to', 'diagnosed by', 'classify', 'screening', 'score'.
(b) Diagnostic Criteria: Phrases such as "meeting X for Y" suggest that X acts as a standard or criterion for diagnosing or defining Y.
(c) Aiding in Identification and Differentiation: Phrases indicating that X can 'help identify', 'distinguish', 'differentiate', 'discriminate', or 'classify' B (a disease), pointing to X's usefulness in differentiating Y from other conditions or in making a definitive diagnosis of Y.
characteristic-of: This relations describe the symptoms, clinical manifestations, or distinct features of a disease. "X is characteristic-of Y" has following types:
(a) Characteristic Identification: Sentences where X is described with phrases like 'characterized by', 'symptom of', 'clinical expression', 'hallmark', signifying Y as a characteristic or symptom of X. For instance, "X is characterized by Y".
(b) Manifestation Descriptions: Sentences where X(disease) is noted to 'present', 'show', 'exhibit', 'report', 'indicate', 'demonstrate', or 'have' a symptom, namely Y. For example, "X presents/shows/exhibits Y".
(c) Marker Identification: Sentences where X is identified by Y, or Y is mentioned as a marker or biomarker for X, or X is more sensitive to Y. For instance, "X is identified by Y" or "Y is a marker/biomarker for X".
(d) Accompaniment Patterns: Sentences where X is usually accompanied by Y. For example, "X with/accompanied by Y".
(e) Quantitative Changes: Sentences indicating that Y is higher, decreased, or increased in X (disease). For instance, "Y is higher/decreased/increased in X".
(f) Observation in Disease: Sentences where Y was observed, found, or occurs in X. For example, "Y was observed/found/occur in X".
associated-with: This relation describes an association or connection between X and Y, where changes in X may correspond with changes Y or changes in Y correspond with X. The associated with relation doesn't have direction such that associated-with(e1,e2) and associated-with(e2,e1) are equivalent. It not necessarily imply a cause-and-effect relation.
treatment-for: This relation type highlights the link between X(treatments or interventions) and their impact on Y(a specific disease or condition). It includes a range of treatment methods like medication, psychotherapy, lifestyle modifications, surgical procedures, and others. "X is treatment-for Y" has following types:
1. Application in Treatment: Phrases like "Use X to treat Y", "X is a treatment/intervention for Y" or "Y treated/medicated with X" indicate the use of X in the treatment of Y.
2. Signs of Improvement: Phrases where X 'improves', 'alleviates', 'suppresses', 'shows efficacy in', 'benefits', 'prevents', or 'reverses' Y, suggesting a beneficial impact of X on Y.
hyponym-of: This relation can indicate a hierarchical link, with X being a subordinate or specific instance of Y, or a constitutive link, showing that X is a component or part of Y. "X is hyponym-of Y" has following types:
(a) Direct Categorization: Sentences where "X is a Y" are used, signifying that X belongs to the category or class of Y.
(b) Appositive Formulation: Sentences in the form of "X, a Y", clearly identifying X as a type or instance of Y.
(c) Indicators of Inclusion or Example: Phrases using 'include', 'involve', 'such as', 'like', 'e.g.', to demonstrate that X is an example or part of the broader category Y.
(d) Alternative Naming: The use of "also called" to connect X with Y, indicating that X is another name or term for Y.
(e) Usage of Colon for Definition: Sentences that use a colon (":") to lead into a description, often employed to list or define hyponyms or specific instances under a broader category.
abbreviation-for: This relation type refers to the use of X(abbreviations or acronyms) to represent Y(longer terms or phrases), especially in medical or technical fields. For example, "AIDS" is an abbreviation for "Acquired Immune Deficiency Syndrome", and "MRI" stands for "Magnetic Resonance Imaging".
occurs-in: The occurs in relation applies to scenarios where X(an event, condition, or phenomenon) is observed in Y(groups or settings). Example: "Condition X occurs in populations with attributes Y".
located-in:"X is located-in Y" occurs in following three scenarios:
(a) Gene or Region Specificity: This relation applies when discussing the location of genes or specific regions. For instance, "Gene/region X is located in gene/region Y".
(b) Disease, Symptom, or Physiological Activity Location: highlights the physical or anatomical location of medical conditions or biological processes. For example, "Disease/symptom/physiological activity X occurs in region Y".
(c) Localization Indicators for Signs: This relation also applies to sentences that indicate where signs of a particular condition (like a disease, gene, region, etc.) are found. For example, "X of Y (region/gene/. . .)".

Figure 7: The Annotation guide of the MDKG dataset.

<think>
Let's analyze the given sentence: "Monitoring prolactinemia and addressing hyperprolactinemia is paramount in women with <e1>schizophrenia</e1> being treated with <e2>prolactin-increasing antipsychotics</e2>."
We need to determine the relationship between <e1>schizophrenia</e1> and <e2>prolactin-increasing antipsychotics</e2>.
Let's break this down step by step.
1. **Identify the entities**:
- <e1>schizophrenia</e1> is the disease or condition.
- <e2>prolactin-increasing antipsychotics</e2> is a type of medication or treatment.
2. **Analyze the relationship**:
- The sentence states that schizophrenia is associated with a specific type of antipsychotics (prolactin-increasing antipsychotics) being used in the treatment or management of schizophrenia.
- We need to determine if this relationship is that schizophrenia leads to or is treated by these antipsychotics, or if it's some other relationship.
3. **Apply the definitions**:
- **treatment-for**: This relationship would apply if the antipsychotics are used to treat or manage schizophrenia. The sentence mentions that these antipsychotics are used specifically in the context of women with schizophrenia, implying a therapeutic use.
- **risk-factor-of**: This would apply if the antipsychotics increased the risk of schizophrenia, but here it's the other way around; the antipsychotics are being used because of schizophrenia, not the other way.
- **associated-with**: While it's possible that there could be an association, the sentence is more about the use of these antipsychotics as a therapeutic intervention rather than just an association.
- **hyponym-of**: This would apply if prolatcin-increasing antipsychotics were a subtype or specific example of a larger category, but here it seems to be a specific type of antipsychotic used for a particular condition.
- **causes**: This wouldn't apply because the sentence suggests the antipsychotics are being used to treat, not cause, schizophrenia. - **characteristic-of**: This would describe a characteristic of the condition, but here we are describing the treatment rather than a characteristic of schizophrenia.
4. **Conclusion**:
- The sentence clearly states that these antipsychotics are being used in the context of women with schizophrenia. This implies that the antipsychotics are a form of treatment specifically for schizophrenia. The key phrase is "being treated with," which indicates that the antipsychotics are used to manage or treat the condition of schizophrenia.
Given that the sentence directly states that antipsychotics are used in the treatment of schizophrenia, this fits squarely into the definition of a treatment relationship. The antipsychotics are specifically used to address or manage schizophrenia, which is the essence of the treatment-for relationship.
Therefore, the relationship here is that the antipsychotics are used to treat schizophrenia.
</think> <answer>treatment-for(e2,e1)</answer>

Figure 8: An example of chain-of-thought (COT) reasoning output from R1-RE-7B

<think> The sentence describes an interplay between DISC1 (intrinsic signaling pathway) and GABA signaling (extrinsic signaling pathway) in the context of neurogenesis, which is affected by susceptibility genes. The susceptibility genes here are DISC1 and GABA. The relation is about how these pathways influence neurogenesis and their impact on neuronal development and risk for mental disorders.
The main focus is on how the presence of DISC1 and GABA signaling, which are considered susceptibility pathways, influence neurogenesis. Since susceptibility pathways are risk factors for certain conditions (in this case, potentially mental disorders), we can infer that DISC1 is a risk-factor-of neurogenesis in the context of developmental processes and experiences.
However, the sentence does not directly indicate that DISC1 is a treatment or method for neurogenesis. Therefore, the relation "risk-factor-of" seems most appropriate here, as it captures the role of DISC1 in influencing the risk or likelihood of neurogenesis being affected by susceptibility genes.
The direction of the relation is from DISC1 to neurogenesis, indicating that the presence of DISC1 (as a susceptibility pathway) impacts neurogenesis. </think>
<answer> risk-factor-of(e1,e2) </answer>

Figure 9: An example of chain-of-thought (COT) reasoning output from Qwen-2.5-7B-Instruct