RADAR: Reasoning as Discrimination with Aligned Representations for LLM-based Knowledge Graph Reasoning

Bo Xue*, **Yuan Jin***, **Luoyi Fu**, **Jiaxin Ding**†, **Xinbing Wang**
School of Computer Science, Shanghai Jiao Tong University
{sappho_x, jiaxinding}@sjtu.edu.cn

Abstract

Knowledge graph reasoning (KGR) infers missing facts, with recent advances increasingly harnessing the semantic priors and reasoning abilities of Large Language Models (LLMs). However, prevailing generative paradigms are prone to memorize surface-level co-occurrences rather than learning genuine relational semantics, limiting out-of-distribution generalization. To address this, we propose RADAR, which reformulates KGR from generative pattern matching to discriminative relational reasoning. We recast KGR as discriminative entity selection, where reinforcement learning enforces relative entity separability beyond token-likelihood imitation. Leveraging this separability, inference operates directly in representation space, ensuring consistency with the discriminative optimization and bypassing the generation-induced hallucinations. Across four benchmarks, RADAR achieves 5–6% relative gains on link prediction and triple classification over strong LLM baselines, while increasing task-relevant mutual information in intermediate representations by 62.9%, indicating more robust and transferable relational reasoning.

1 Introduction

Knowledge graphs (KGs) represent facts as structured triples and support applications such as semantic search and question answering (Chen et al., 2020; Huang et al., 2019; Wang et al., 2024). However, real-world KGs are inevitably incomplete, limiting their effectiveness in downstream use (Shen et al., 2022b). Knowledge graph reasoning (KGR) addresses this by inferring missing facts from observed triples. Classical embedding-based methods exploit graph structure effectively, but often struggle under sparsity and have limited capacity to express fine-grained relational semantics beyond local topology (Pujara et al., 2017; Yao et al., 2019). Consequently, recent work has shifted toward large language models (LLMs), harnessing their rich semantic priors and reasoning capabilities to mitigate these limitations (Zhang et al., 2024b; Yao et al., 2025; Li et al., 2024b,a).

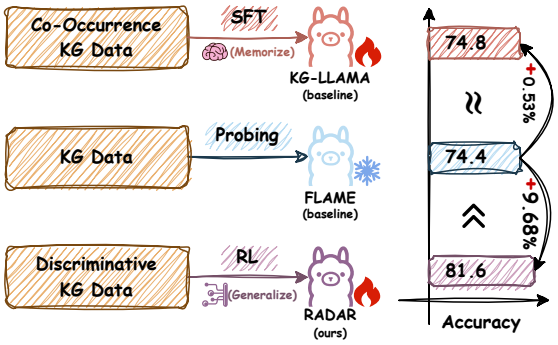

Figure 1: FB15K-237N triple classification with a shared LLaMA backbone: co-occurrence-based SFT yields only marginal gains over the frozen baseline, whereas RADAR substantially improves accuracy by optimizing discriminative relational reasoning.

A central requirement of KGR is *generalization*: the ability to answer queries involving entity–relation combinations unseen during training (Liu et al., 2021; Bordes et al., 2013). Yet prevailing LLM-based paradigms formulate KGR as sequence modeling, optimizing next-token likelihood over serialized textual triples. This formulation encourages a shortcut: models minimize loss by exploiting surface co-occurrences between entity names and relations rather than learning relation-conditioned validity (Zhang et al., 2024a; Kang and Choi, 2023; Ju et al., 2024). Such behavior yields strong in-distribution performance but fails under distribution shift where those co-occurrence statistics no longer hold. Crucially, this reliance on co-occurrence shortcuts is further exacerbated by supervised fine-tuning (SFT), as cross-entropy objectives reinforce token-level imitation over relational reasoning (Chu et al., 2025; Lv et al., 2025). As illustrated in Figure 1, co-occurrence-based SFT yields only marginal gains over the frozen baseline,

*Equal contribution. † Corresponding author.

Findings of the Association for Computational Linguistics: EMNLP 2026, pages 1209–1223
October 24-29, 2026 ©2026 Association for Computational Linguistics

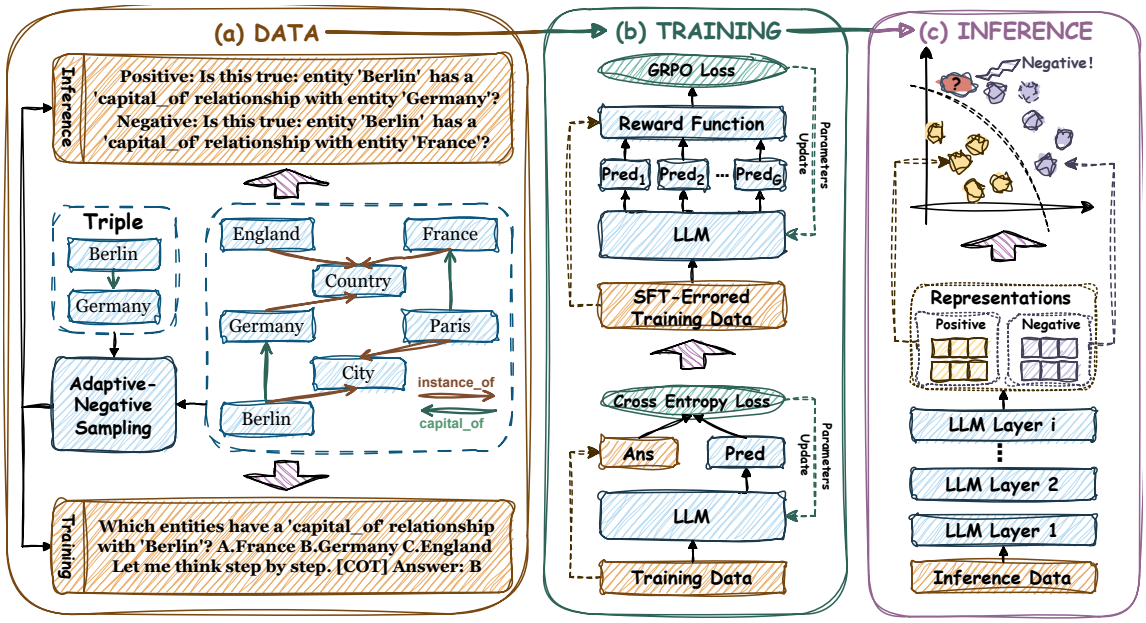

Figure 2: RADAR, a data-to-inference alignment framework for discriminative KGR.

highlighting the necessity for a paradigm shift from generative pattern matching to relational reasoning.

To address co-occurrence-driven shortcuts, we propose RADAR (Figure 2), which fundamentally shifts KGR from generative pattern matching to discriminative relational reasoning. By recasting KGR as discriminative entity selection, we redirect the learning signal from token-level imitation to the relative separability of ground truth against hard distractors, rendering co-occurrence shortcuts insufficient (Si et al., 2022). This reformulation pivots KGR from local token-level prediction to global entity discrimination within a discrete candidate space, where Reinforcement Learning (RL) naturally aligns with optimizing global discrete outcomes and effectively resolving the objective mismatch inherent in token-level SFT (Chu et al., 2025). Operationalizing this shift necessitates staged training, where RL exploits SFT-grounded structure to amplify reward-driven contrast between correct entities and distractors, transforming the objective from likelihood-based imitation to generalizable discrimination (Lv et al., 2025). As a consequence of optimizing for relative entity separability, relational information is encoded in discriminative latent representations rather than through autoregressive decoding (Zou et al., 2023). Accordingly, we perform inference directly in representation space, aligning inference with the discriminative signals induced during training and avoiding generation-induced hallucinations.

Across four KGR benchmarks, RADAR consistently achieves an average 5–6% relative improvement over strong LLM-based baselines on link prediction and triple classification. Ablations show that the gains come from a tightly integrated design that aligns what the model is trained to optimize, what it is allowed to rely on, and what is ultimately extracted for reasoning. Furthermore, our proposed information-theoretic probe reveals that RADAR increases task-relevant mutual information in intermediate representations by 62.9% on average, consistent with improved inductive robustness and domain transfer. Our contributions are threefold:

- We expose co-occurrence shortcuts as a fundamental bottleneck in LLM-based KGR and recast KGR as a discriminative relational reasoning problem, moving beyond autoregressive pattern matching to achieve robust out-of-distribution generalization.
- We introduce RADAR, a data-to-inference alignment framework that restructures supervision, optimization, and inference around discrete entity separation, inducing relational semantics to emerge as discriminative structures within the representation space and bypassing surface-level co-occurrence shortcuts.
- Extensive evaluations across four benchmarks demonstrate superior performance over strong baselines, while information-theoretic quantification validates that RADAR substantially enhances task-relevant mutual information to achieve robust inductive generalization.

2 Methods

2.1 Setup

A knowledge graph $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{T}\}$ is a collection of an entity set $\mathcal{E}$, a relation set $\mathcal{R}$, and a triple set $\mathcal{T}$. Each triple is denoted as $(h, r, t) \in \mathcal{T}$, where $h, t \in \mathcal{E}$ represent the entities and $r \in \mathcal{R}$ denotes the relation. Let $\ell(\cdot)$ denote the function that maps each entity or relation to its textual sequence for LLM processing.

In this work, we focus on two standard KGR tasks: triple classification and link prediction. Triple classification determines whether a triple (h, r, t) is valid. Link prediction aims to identify the missing entity in an incomplete triple $(h, r, ?)$.

2.2 Task Reformulation for Discriminative Selection

Prevailing generative paradigms (Saxena et al., 2022; Yao et al., 2025) optimizing serialized triples (e.g., "$\ell(h)$ $\ell(r)$ $\ell(t)$") inherently biases models toward minimizing loss via surface-level entity-relation co-occurrences rather than relational validity (Kang and Choi, 2023; Ju et al., 2024). To redirect this bias, we recast KGR as discriminative entity selection within a constrained candidate space. This reformulation fundamentally alters the learning signal: the model is compelled to explicitly discriminate ground truth from distractors, thereby forcing the optimization trajectory toward inducing discriminative latent structures governed by relational validity, rendering co-occurrence shortcuts insufficient for objective minimization.

Formally, given a query triple $(h, r, ?)$, we define $\mathcal{E}_{\text{gt}}(h, r) = \{t \mid (h, r, t) \in \mathcal{T}\}$ as the ground-truth tail entity set. We construct a candidate set $\mathcal{C}(h, r) = \{c_1, c_2, \ldots, c_K\}$, where $\mathcal{C}(h, r) = \mathcal{E}_{pos}(h, r) \cup \mathcal{E}_{neg}(h, r)$, $\mathcal{E}_{pos}(h, r) \subseteq \mathcal{E}_{\text{gt}}(h, r)$ denotes positive entities and $\mathcal{E}_{neg}(h, r) \subset \mathcal{E} \backslash \mathcal{E}_{\text{gt}}(h, r)$ denotes negative entities. Each training instance is formatted as shown in Figure 3.

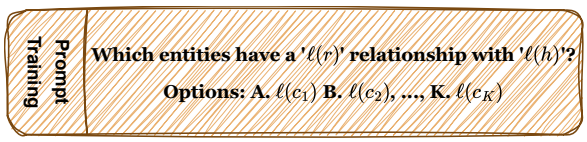

Figure 3: Training prompt.

The model is trained to output the option labels corresponding to entities in $\mathcal{E}_{pos}(h, r)$.

Hierarchical Task Difficulty. To progressively challenge the model's relational reasoning capabilities, we design a hierarchical framework along two dimensions: answer cardinality and negative sample hardness.

For answer cardinality, we construct two task variants: (1) *single-answer*: $|\mathcal{E}_{pos}(h, r)| = 1$; (2) *variable-answer*: $|\mathcal{E}_{pos}(h, r)| \geq 1$, where the number of correct entities is not given a priori, encompassing both one-to-one and one-to-many relations and requiring cardinality determination.

For negative sample hardness, we stratify negative candidates based on their plausibility given the query. We use a pre-trained knowledge graph embedding (KGE) model to score all potential entities and partition $\mathcal{E}_{neg}(h, r)$ into three tiers based on plausibility. (1) *Tier 1 (Easy)*: lowest-scoring entities that are structurally implausible as tails for the query; (2) *Tier 2 (Medium)*: randomly sampled entities; (3) *Tier 3 (Hard)*: highest-scoring entities that are highly confusable with ground-truth answers, requiring fine-grained relational reasoning.

2.3 Optimization for Discriminative Relational Reasoning

To overcome the pattern matching, we develop a two-stage training paradigm tightly coupled with discrete entity selection formulation. While SFT grounds output structure, its cross-entropy objective inherently favors token-level imitation over relational reasoning (Lv et al., 2025). This mismatch necessitates RL, which pivots the learning signal from token likelihood to global outcome validity (Chu et al., 2025). Crucially, the reformulation in Section 2.2 yields explicit, verifiable rewards, enabling stable RL and promoting genuine relational discrimination over statistical mimicry.

Stage I: Supervised Fine-Tuning. We construct chain-of-thought (CoT) reasoning traces (Wei et al., 2022) for each training instance, where each trace verifies semantic compatibility of candidate entities with the query $(h, r, ?)$ before producing the final answer set (see Appendix E for details). Given a training dataset $\mathcal{D}_{\text{train}}$, where each instance consists of a discriminative prompt x and a target sequence comprising the CoT reasoning and the final answer. The model is fine-tuned using the standard next-token prediction objective.

Stage II: Reinforcement Learning. We apply Group Relative Policy Optimization (GRPO) (Shao et al., 2024), leveraging group-normalized advantages for stable policy updates, to enhance generalization on challenging entity-relation compositions.

Let $\hat{y}$ denote the model output for input x, and let $\hat{\mathcal{A}}(\hat{y}) \subseteq \mathcal{C}(h, r)$ denote the set of entities extracted from $\hat{y}$. To amplify reasoning signals on instances where SFT fails, we construct an error-focused dataset $\mathcal{D}_{\text{error}}$ by evaluating the SFT model on $\mathcal{D}_{\text{train}}$ and retaining only instances where $\hat{\mathcal{A}}(\hat{y}) \neq \mathcal{E}_{pos}(h, r)$.

To guide learning on $\mathcal{D}_{\text{error}}$ while accommodating variable-cardinality predictions, we define a composite reward $R(x, \hat{y})$:

$$R(x, \hat{y}) = \alpha \cdot R_{\text{fmt}}(\hat{y}) + (1 - \alpha) \cdot R_{\text{acc}}(x, \hat{y}) \quad (1)$$

where α is the weighting coefficient, $R_{\text{fmt}}(\hat{y}) \in \{0, 1\}$ verifies format adherence, and $R_{\text{acc}}(x, \hat{y})$ measures answer accuracy. For each query $(h, r, ?)$ encoded in x, the accuracy reward is:

$$R_{\text{acc}}(x, \hat{y}) = \frac{2 \cdot |\hat{\mathcal{A}}(\hat{y}) \cap \mathcal{E}_{pos}(h, r)|}{|\hat{\mathcal{A}}(\hat{y})| + |\mathcal{E}_{pos}(h, r)|} \quad (2)$$

This F1-based formulation provides credit assignment for partially correct predictions, which is essential for variable-answer instances. We optimize the policy to maximize the expected reward using GRPO, with details provided in Appendix E.

2.4 Representation-based Inference

By explicitly discriminating against distractors, the task reformulation and staged training shape the model's internal representations toward relational separability. Autoregressive decoding evaluates relational plausibility indirectly through next-token likelihood, introducing a mismatch between token-level generation and the entity-level discriminative separability induced during training (Zou et al., 2023; Orgad et al., 2024). We therefore perform inference directly in representation space, enabling faithful extraction of discriminative relational knowledge and ensuring that the inference mechanism is intrinsically aligned with the discriminative learning signals optimized during training.

For triple classification, we extract the model's internal representations to assess triple plausibility. We construct prompt $\text{PT}_{\text{cls}}(h, r, t)$ as follows:

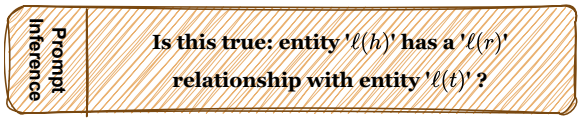

Figure 4: Inference prompt.

We extract the hidden state at the final token position from an intermediate layer l, as intermediate representations have been shown to better preserve factual knowledge than final-layer outputs (Skean et al., 2025):

$$\mathbf{z}^{(l)}(h, r, t) = (\text{LLM}^{(l)}(\text{PT}_{\text{cls}}(h, r, t)))_T, \quad (3)$$

where $\mathbf{z}^{(l)}(h, r, t) \in \mathbb{R}^d$ denotes the extracted hidden state, $\text{LLM}^{(l)}(\cdot)$ represents output of the first l layers of the fine-tuned model, $(\cdot)_T$ extracts the representation at position T (the final token), and d is the hidden dimension.

We then train a binary classifier $f_\phi : \mathbb{R}^d \to [0, 1]$ that maps the extracted representation to a plausibility score $s(h, r, t) = f_\phi(\mathbf{z}^{(l)}(h, r, t))$. Training data consists of positive samples from ground-truth triples and negative samples constructed by corrupting tail entities (Bordes et al., 2013). We optimize f_ϕ using binary cross-entropy loss and implement f_ϕ as a two-layer MLP, which maps $\mathbf{z}^{(l)}$ through a hidden layer of dimension d_v to produce the binary classification score.

For link prediction, where the goal is to identify the missing tail entity in $(h, r, ?)$, we adopt a retrieval-then-reranking approach (Li et al., 2025). Given the large entity set size, exhaustively scoring all entities is computationally prohibitive. We therefore leverage a lightweight pre-trained KGE model to retrieve the top-n structurally plausible candidates ($n \ll |\mathcal{E}|$), then rerank them using our learned classifier f_ϕ. For each retrieved candidate e_i, we compute $s(h, r, e_i) = f_\phi(\mathbf{z}^{(l)}(h, r, e_i))$ and rank candidates by this score.

2.5 Task-Adaptive Information Quantification

To quantify discriminative relational signals internalized during our reformulated KGR training, we propose a task-adaptive mutual information measure defined over internal representations $\mathbf{z}^{(l)}$. While standard sliced mutual information (SMI) relies on random projections that treat all directions equally, this indiscriminate averaging can dilute task-relevant information in LLM representations (Goldfeld and Greenewald, 2021; Wongso et al., 2023). In contrast, we derive projections from the learned probing classifier f_ϕ to target the most discriminative subspace, concentrating information estimation on directions that encode relational separability induced by discriminative KGR training.

Since f_ϕ is optimized to distinguish valid triples, its weight matrix encodes discriminative directions in the representation space. For N samples with representations $\mathbf{Z}^{(l)} \in \mathbb{R}^{d \times N}$, we define task-adaptive projections using the first layer of f_ϕ:

$$\mathbf{V} = \text{PReLU}(\mathbf{W}_1 \mathbf{Z}^{(l)} + \mathbf{b}_1) \in \mathbb{R}^{d_v \times N}, \quad (4)$$

where $\mathbf{W}_1 \in \mathbb{R}^{d_v \times d}$ and $\mathbf{b}_1 \in \mathbb{R}^{d_v}$ are the trained parameters. Each row $\mathbf{v}_i \in \mathbb{R}^N$ of $\mathbf{V}$ corresponds to the representations projected onto a dimension optimized for KGR.

Let $Y \in \{0, 1\}^N$ denote the ground-truth validity labels. We then quantify the task-specific information encoded in the representation space by averaging the mutual information between the projected feature activations and the labels:

$$\mathcal{I}_{\text{task}} = \frac{1}{d_v} \sum_{i=1}^{d_v} \hat{I}(\mathbf{v}_i; Y), \quad (5)$$

where $\hat{I}$ is mutual information computed via the KSG method (Kraskov et al., 2004). This metric quantifies the KGR-relevant information within the representations that is captured by the discriminative features of f_ϕ, enabling comparison across training strategies (details in Appendix D).

3 Experiment

Our experiments address the following research questions: **RQ1:** Does RADAR achieve strong and consistent performance across standard KGR benchmarks and tasks? **RQ2:** How do the core design components of RADAR individually and synergistically contribute to performance and generalization? **RQ3:** Can RADAR achieve robust inductive generalization to unseen entities and transfer relational knowledge to domain-related tasks?

3.1 Experiment Settings

Datasets. We evaluate RADAR on four benchmarks across two tasks. For link prediction, we use FB15K-237 (Schlichtkrull et al., 2018) and FB15K-237N (Lv et al., 2022). For triple classification, we employ WN18RR (Dettmers et al., 2018), UMLS (Yao et al., 2019), and FB15K-237N. Dataset details are provided in Appendix B.

Baselines. We compare RADAR against two categories of baselines: 1) Knowledge Graph Embedding models, which map entities and relations into low-dimensional vector spaces, including TransE (Bordes et al., 2013), DistMult (Yang et al., 2014), ComplEx (Trouillon et al., 2016), RotatE (Sun et al., 2019), ConvE (Dettmers et al., 2018), and TuckER (Balažević et al., 2019). 2) Language Model-based methods, which leverage the textual semantics of KGs. These include KG-BERT (Yao et al., 2019), MTL-KGC (Kim et al., 2020), KG-S2S (Chen et al., 2022a), SimKGC (Wang et al., 2022), iGT (Luo et al., 2025), CSPromp-KG (Chen et al., 2023), COSIGN (Li et al., 2024b), KG-LLAMA (Yao et al., 2025), CD (Li et al., 2024a), LLAMA-ICL, Structure IT, KoPA (Zhang et al., 2024b) and FLAME (Xue et al., 2024). Among these, KG-LLAMA and FLAME are the most directly comparable baselines because they rely solely on LLaMA's intrinsic capabilities without external structural embeddings, in contrast to hybrid approaches such as KoPA. Further details on these baselines are provided in Section A.

Evaluation Metrics. For link prediction, we report the Mean Reciprocal Rank (MRR) and Hits@k ($k = 1, 3, 10$) following standard protocols.

Implementation Details. Unless otherwise noted, we adopt the variable-answer setting with Tier 2 negative sampling (random negatives) for all experiments. To ensure fair comparison with KG-LLAMA and FLAME, we adopt LLaMA (Touvron et al., 2023) as the backbone model across all configurations. For link prediction inference and negative sample stratification in data construction (Section 2.2), we employ TuckER (Balažević et al., 2019) as the KGE model to score candidate entities. Additional implementation details are provided in Appendix C.

3.2 Main Results (RQ1)

Table 1 reports link prediction results of RADAR on FB15K-237 and FB15K-237N. Specifically, RADAR attains the best MRR across both datasets, outperforming all KGE-based and language model-based baselines, while ranking first or second on all Hits@k metrics. On FB15K-237, RADAR reaches an MRR of 0.377, corresponding to a 2.4% relative improvement over the previous best model COSIGN. Notably, RADAR demonstrates substantial improvements on the more challenging FB15K-237N dataset, achieving a 5.3% improvement in MRR over COSIGN.

To complement link prediction, we further evaluate triple classification on WN18RR, FB15K-237N, and UMLS as shown in Table 2. RADAR achieves the best accuracy on WN18RR and FB15K-237N,

Table 1: Link prediction performance (MRR and Hits@k) on FB15K-237 and FB15K-237N. Best scores are in **bold** and second-best are underlined.

Paradigm	Model	FB15K-237				FB15K-237N			
		MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
Embedding -based	TransE	0.279	0.198	0.376	0.441	0.255	0.152	0.301	0.459
	DistMult	0.281	0.199	0.301	0.446	0.209	0.143	0.234	0.330
	ComplEx	0.278	0.194	0.297	0.450	0.249	0.180	0.276	0.380
	RotatE	0.338	0.241	0.375	0.533	0.279	0.177	0.320	0.481
	ConvE	0.312	0.225	0.341	0.497	0.273	0.192	0.305	0.429
	TuckER	0.347	0.253	0.382	0.536	0.301	0.217	0.332	0.463
Language Model -based	KG-BERT	0.237	0.169	0.260	0.427	0.203	0.139	0.201	0.403
	MTL-KGC	0.267	0.172	0.298	0.458	0.241	0.160	0.284	0.430
	SimKGC	0.333	0.246	0.362	0.510	0.372	0.289	0.402	0.534
	KG-S2S	0.336	0.257	0.373	0.498	0.353	0.282	0.385	0.495
	CSPromp-KG	0.358	0.269	0.393	0.538	0.360	0.281	0.395	0.511
	CD	–	–	–	–	0.372	0.288	0.410	0.530
	COSIGN	0.368	**0.315**	**0.434**	0.520	0.394	**0.355**	0.457	0.526
	KG-LLAMA	0.238	0.165	0.272	0.423	–	–	–	–
RADAR		**0.377**	0.273	0.421	**0.579**	**0.415**	0.301	**0.476**	**0.633**

Table 2: Triple classification accuracy on WN18RR, FB15K-237N, and UMLS. Best scores are in **bold** and second-best are underlined.

Paradigm	Model	WN18RR	FB15K-237N	UMLS
Embedding -based	TransE	88.4	69.7	84.5
	DistMult	85.1	58.7	86.4
	ComplEx	84.1	65.7	87.1
	RotatE	88.2	68.5	87.7
Language Model -based	KG-BERT	91.6	56.0	77.3
	LLAMA-ICL	50.2	59.2	55.5
	KG-LLAMA	92.1	74.8	85.8
	Structure IT	92.7	76.4	89.9
	FLAME	93.8	74.4	86.6
	KoPA	94.9	77.7	**92.6**
RADAR		**95.3**	**81.6**	91.7

and is second on UMLS (91.7%), slightly below KoPA (92.6%). KoPA leverages external structural embeddings, whereas RADAR operates in the LLM-only setting. Focusing on the LLM-only setting with the same LLaMA backbone, RADAR substantially improves over KG-LLAMA and FLAME, yielding an average relative gain of 6.1% and 5.7% across the three datasets, respectively, suggesting reduced reliance on entity–relation co-occurrence shortcuts and stronger generalization in LLaMA-based KGR.

Finally, Table 3 studies task variants on FB15K-237N by varying answer cardinality and negative hardness. The variable-answer setting performs better than single-answer, suggesting that exposing multiple correct tails for one-to-many relations provides richer supervision for relation-conditioned discrimination. Under the variable-answer setting,

Table 3: Triple classification accuracy on FB15K-237N under controlled ablations of task difficulty: answer cardinality and negative hardness.

Answer Cardinality *(Fixing Negatives: Tier 2)*		**Negative Hardness** *(Fixing Cardinality: Variable)*	
Setting	Accuracy	Setting	Accuracy
Single-answer	79.9	Tier 1 *(Easy)*	80.2
Variable-answer	81.6	Tier 2 *(Medium)*	81.6
		Tier 3 *(Hard)*	81.0

Tier-2 is marginally better than Tier-3 and Tier 1, and overall performance differences across tiers remain small, motivating our default choice of Tier-2 in subsequent experiments. See Appendix F for further analysis on negative hardness.

Taken together, these results show that RADAR delivers consistent gains across link prediction and triple classification under a unified discriminative training-and-inference pipeline, outperforming prior LLM-only baselines while remaining competitive against methods that explicitly fuse KG structural embeddings.

3.3 Ablation Results (RQ2)

We ablate RADAR along three axes—data formulation (serialization vs. discrimination), training objective (sft vs. two-stage training), and inference (generation vs. representation extraction)—using four variants named by DATA–TRAIN–INFER. *Serial-SFT-Generation* follows the KG-LLaMA-style generative setup, fine-tuning on serialized triples (e.g., Is this true: $\ell(h)\ \ell(r)\ \ell(t)$?). *Serial-SFT-Extraction* keeps the same serialized SFT

setup but replaces decoding with representation extraction-based inference (Section 2.4), isolating the effect of inference. *Discriminative-SFT-Extraction* replaces serialization with the discriminative candidate-selection formulation (Section 2.2), which reduces entity–relation co-occurrence shortcuts while keeping SFT-only training and extraction inference, isolating the impact of task formulation. *Discriminative-Full-Extraction* represents our full framework with discriminative formulation, two-stage full training (Section 2.3), and extraction inference, isolating the effect of two-stage training.

Table 4: Ablation across task formulation, training objective, and inference strategy on triple classification accuracy (WN18RR, FB15k-237N, UMLS).

Method	WN18RR	15K-237N	UMLS
Serial-SFT-Generation	0.921	0.748	0.858
Serial-SFT-Extraction	0.925	0.753	0.861
Discriminative-SFT-Extraction	0.945	0.793	0.898
Discriminative-Full-Extraction	**0.953**	**0.816**	**0.917**

Table 4 quantifies the contribution of each design axis in RADAR. First, replacing autoregressive decoding with representation-based inference (*Serial-SFT-Extraction*) yields consistent gains over generation (*Serial-SFT-Generation*), suggesting that internal representations capture relational information more reliably than autoregressive outputs. Second, reformulating KGR as a discriminative task leads to substantial performance improvements. Specifically, *Discriminative-SFT-Extraction* improves over *Serial-SFT-Extraction* by an average of 3.9% across benchmarks. These improvements align with our hypothesis that the discriminative formulation mitigates entity–relation co-occurrence shortcuts and promotes reasoning over relational semantics. Incorporating RL refinement yields further gains, with *Discriminative-Full-Extraction* achieving an average relative improvement of 6.5% over the standard serial SFT generation baseline. Together, these results show that effective relational reasoning in LLM-based KGR requires a tightly integrated design across task formulation, optimization, and inference. In this system, outcome-based optimization does not merely refine supervised learning, but works in concert with the discriminative setup and representation-based inference to resolve non-trivial challenges in aligning what the model is trained to optimize, what it is allowed to rely on, and what knowledge is ultimately extracted for reasoning.

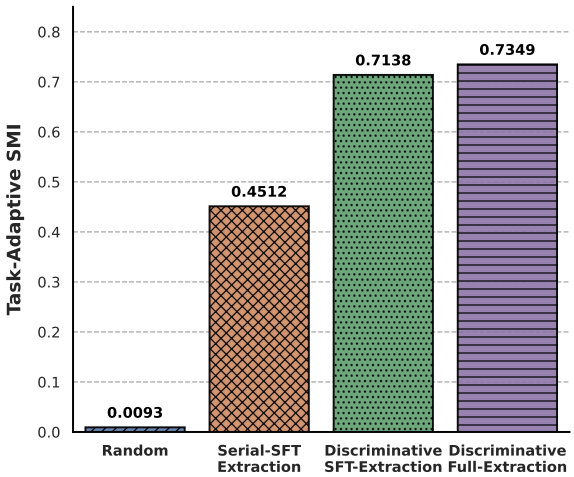

Figure 5: Task-Adaptive SMI of intermediate representations across different methods on WN18RR.

To better interpret the performance gains observed in the ablation study, we measure how each design choice affects the amount of KGR-relevant information encoded in intermediate representations using task-adaptive mutual information $\mathcal{I}_{\text{task}}$. As shown in Figure 7, RADAR yields a 62.9% average relative increase in $\mathcal{I}_{\text{task}}$ over the serialized SFT baseline on WN18RR (please refer to Appendix D for corresponding analyses on other datasets). Discriminative reformulation establishes a representation space suitable for relational discrimination, while reinforcement learning further aligns this space with outcome-level relational correctness beyond supervised likelihood. Together, these results indicate that the gains of RADAR stem from systematic changes in how task-relevant information is organized in the representation space.

Table 5: Link prediction performance on FB15K-237N across different LLM backbones.

Method	MRR	H@1	H@3	H@10
LLaMA-7B (Touvron et al., 2023)				
Serial-SFT-Extraction	0.365	0.229	0.438	0.584
Discriminative-Full-Extraction	**0.415**	**0.301**	**0.476**	**0.633**
Pythia-6.9B (Biderman et al., 2023)				
Serial-SFT-Extraction	0.353	0.259	0.401	0.616
Discriminative-Full-Extraction	**0.384**	**0.267**	**0.437**	**0.635**
Qwen3-8B (Yang et al., 2025)				
Serial-SFT-Extraction	0.402	0.287	0.442	0.622
Discriminative-Full-Extraction	**0.426**	**0.312**	**0.499**	**0.641**

We evaluate the architectural robustness of RADAR by examining whether its relative improvements are preserved across diverse LLM backbones. As shown in Table 5, RADAR consistently outperforms the corresponding Serial-SFT-Extraction baselines across all evaluated architectures. Specifically, on Qwen3-8B, RADAR yields an relative improvement of 6.0% in MRR and 8.7% in Hits@1 over the Serial-SFT-Extraction baseline. These results indicate that RADAR is robust to backbone choice, yielding consistent performance improvements across diverse LLM architectures.

3.4 Additional Results (RQ3)

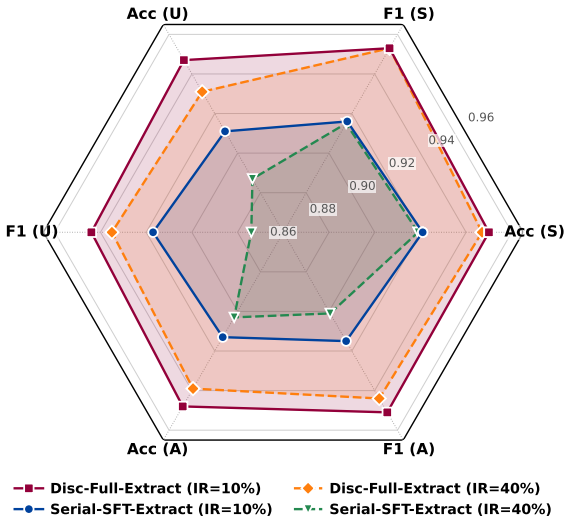

Figure 6: Inductive generalization performance for triple classification on WN18RR under varying inductive rates (IR). We report accuracy on seen (S), unseen (U), and overall (A) test triples.

We evaluate inductive generalization under an entity-disjoint setting in which a subset of entities is completely excluded from training, directly testing the model's ability to reason beyond observed entity identities (Teru et al., 2020). Following (Chen et al., 2022b), we define an inductive rate (IR) as the proportion of entities excluded from training. Given IR=ρ, a fraction ρ of entities is designated as inductive, and all training triples involving them are removed. We stratify test triples into seen (S) where all entities appeared in the training set, unseen (U) involving at least one inductive entity, and (A) denoting the aggregate set. This setting presents a stringent generalization challenge, as inductive entities are never observed during training.

Discriminative-Full-Extraction consistently outperforms the Serial-SFT baseline across all inductive rates, with particularly pronounced gains on unseen triples. At IR=40%, the baseline exhibits a sharp degradation on unseen triples, while our method remains substantially more stable. This gap supports our hypothesis that serialization-based SFT encourages reliance on entity co-occurrence statistics, which fails when such correlations are unavailable. In contrast, RADAR encourages the model to internalize relation-conditioned distinction patterns that are decoupled from specific entity identities. As a result, the learned relational criteria can be consistently applied to unseen entities at inference time, indicating that the components interact in a principled and complementary manner rather than functioning as isolated heuristics.

To evaluate whether knowledge learned through KGR generalizes beyond the training task, we conduct zero-shot evaluations on MMLU (Hendrycks et al., 2020) subjects aligned with the UMLS domain, following the protocol of (Zhang et al., 2024b). As shown in Table 6, KGR training on UMLS leads to an average performance gain of 11.6% across five subjects. These gains suggest that our KGR training encourages relation-conditioned and compositional reasoning that generalizes to domain-adjacent tasks, rather than reliance on memorized triples.

Table 6: Zero-shot performance on domain-aligned MMLU subjects after UMLS-based KGR training.

Subjects	w/o Training	w/ Training
Clinical	0.340	**0.455**
College Medicine	0.341	**0.397**
High School Biology	0.313	**0.439**
High School Chemistry	0.271	**0.355**
Medical Genetics	0.290	**0.490**

4 Conclusion

In this work, we introduce RADAR, a framework that fundamentally realigns LLM-based KGR from autoregressive pattern matching to discriminative relational reasoning. By unifying task reformulation, optimization, and inference toward discriminative entity selection, our approach suppresses memorization of surface statistics in favor of robust relational semantics. Validated by information-theoretic analysis and extensive benchmarks, RADAR not only achieves superior generalization but also establishes a principled route to more reliable and generalizable KGR with LLMs.

5 Limitations

To balance large-scale inference efficiency, RADAR adopts a retrieve-then-rerank paradigm, introducing an inherent recall bottleneck. Additionally, our discriminative alignment via reinforcement learning incurs higher training costs than standard fine-tuning. Future work will focus on optimizing training efficiency and exploring retriever-free LLM reasoning mechanisms.

6 Acknowledgments

This work is funded by NSFC (No. 62525209, T2421002, 623B2071, 92579104, T2542021), National Key Laboratory of Data Space Technology and System and ZTE Industry-University-Research Pre-research Fund.

References

Ivana Balažević, Carl Allen, and Timothy Hospedales. 2019. Tucker: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 5185–5194.

Stella Biderman, Hailey Schoelkopf, Quentin Gregory Anthony, Herbie Bradley, Kyle O'Brien, Eric Hallahan, Mohammad Aflah Khan, Shivanshu Purohit, USVSN Sai Prashanth, Edward Raff, and 1 others. 2023. Pythia: A suite for analyzing large language models across training and scaling. In *International Conference on Machine Learning*, pages 2397–2430. PMLR.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Chen Chen, Yufei Wang, Bing Li, and Kwok-Yan Lam. 2022a. Knowledge is flat: A seq2seq generative framework for various knowledge graph completion. *arXiv preprint arXiv:2209.07299*.

Chen Chen, Yufei Wang, Aixin Sun, Bing Li, and Kwok-Yan Lam. 2023. Dipping plms sauce: Bridging structure and text for effective knowledge graph completion via conditional soft prompting. *arXiv preprint arXiv:2307.01709*.

Mingyang Chen, Wen Zhang, Yushan Zhu, Hongting Zhou, Zonggang Yuan, Changliang Xu, and Huajun Chen. 2022b. Meta-knowledge transfer for inductive knowledge graph embedding. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 927–937.

Zhe Chen, Yuehan Wang, Bin Zhao, Jing Cheng, Xin Zhao, and Zongtao Duan. 2020. Knowledge graph completion: A review. *Ieee Access*, 8:192435–192456.

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*.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *AAAI*.

Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. 2020. Transformer feed-forward layers are key-value memories. *arXiv preprint arXiv:2012.14913*.

Ziv Goldfeld and Kristjan Greenewald. 2021. Sliced mutual information: A scalable measure of statistical dependence. *NeurIPS*, 34:17567–17578.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, and 1 others. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

Xiao Huang, Jingyuan Zhang, Dingcheng Li, and Ping Li. 2019. Knowledge graph embedding based question answering. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 105–113.

Che Jiang, Biqing Qi, Xiangyu Hong, Dayuan Fu, Yang Cheng, Fandong Meng, Mo Yu, Bowen Zhou, and Jie Zhou. 2024. On large language models' hallucination with regard to known facts. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1041–1053.

Tianjie Ju, Yijin Chen, Xinwei Yuan, Zhuosheng Zhang, Wei Du, Yubin Zheng, and Gongshen Liu. 2024. Investigating multi-hop factual shortcuts in knowledge editing of large language models. *arXiv preprint arXiv:2402.11900*.

Cheongwoong Kang and Jaesik Choi. 2023. Impact of co-occurrence on factual knowledge of large language models. *arXiv preprint arXiv:2310.08256*.

Bosung Kim, Taesuk Hong, Youngjoong Ko, and Jungyun Seo. 2020. Multi-task learning for knowledge graph completion with pre-trained language models. In *Proceedings of the 28th international conference on computational linguistics*, pages 1737–1743.

Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. 2004. Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 69(6):066138.

Dawei Li, Zhen Tan, Tianlong Chen, and Huan Liu. 2024a. Contextualization distillation from large language model for knowledge graph completion. *arXiv preprint arXiv:2402.01729*.

Jinpeng Li, Hang Yu, Xiangfeng Luo, and Qian Liu. 2024b. Cosign: Contextual facts guided generation for knowledge graph completion. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 1669–1682.

Muzhi Li, Cehao Yang, Chengjin Xu, Xuhui Jiang, Yiyan Qi, Jian Guo, Ho-fung Leung, and Irwin King. 2025. Retrieval, reasoning, re-ranking: A context-enriched framework for knowledge graph completion. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4349–4363.

Ye Liu, Yao Wan, Lifang He, Hao Peng, and Philip S Yu. 2021. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6418–6425.

Kangyang Luo, Yuzhuo Bai, Cheng Gao, Shuzheng Si, Zhu Liu, Yingli Shen, Zhitong Wang, Cunliang Kong, Wenhao Li, Yufei Huang, and 1 others. 2025. Gltw: Joint improved graph transformer and llm via three-word language for knowledge graph completion. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 11328–11344.

Xin Lv, Yankai Lin, Yixin Cao, Lei Hou, Juanzi Li, Zhiyuan Liu, Peng Li, and Jie Zhou. 2022. Do pre-trained models benefit knowledge graph completion? a reliable evaluation and a reasonable approach. Association for Computational Linguistics.

Xingtai Lv, Yuxin Zuo, Youbang Sun, Hongyi Liu, Yuntian Wei, Zhekai Chen, Lixuan He, Xuekai Zhu, Kaiyan Zhang, Bingning Wang, and 1 others. 2025. Towards a unified view of large language model post-training. *arXiv preprint arXiv:2509.04419*.

Hadas Orgad, Michael Toker, Zorik Gekhman, Roi Reichart, Idan Szpektor, Hadas Kotek, and Yonatan Belinkov. 2024. Llms know more than they show: On the intrinsic representation of llm hallucinations. *arXiv preprint arXiv:2410.02707*.

Jay Pujara, Eriq Augustine, and Lise Getoor. 2017. Sparsity and noise: Where knowledge graph embeddings fall short. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 1751–1756.

Apoorv Saxena, Adrian Kochsiek, and Rainer Gemulla. 2022. Sequence-to-sequence knowledge graph completion and question answering. *arXiv preprint arXiv:2203.10321*.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

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

Jianhao Shen, Chenguang Wang, Linyuan Gong, and Dawn Song. 2022a. Joint language semantic and structure embedding for knowledge graph completion. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1965–1978.

Tong Shen, Fu Zhang, and Jingwei Cheng. 2022b. A comprehensive overview of knowledge graph completion. *Knowledge-Based Systems*, 255:109597.

Qingyi Si, Fandong Meng, Mingyu Zheng, Zheng Lin, Yuanxin Liu, Peng Fu, Yanan Cao, Weiping Wang, and Jie Zhou. 2022. Language prior is not the only shortcut: A benchmark for shortcut learning in vqa. *arXiv preprint arXiv:2210.04692*.

Oscar Skean, Md Rifat Arefin, Dan Zhao, Niket Patel, Jalal Naghiyev, Yann LeCun, and Ravid Shwartz-Ziv. 2025. Layer by layer: Uncovering hidden representations in language models. *arXiv preprint arXiv:2502.02013*.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*.

Xiaojuan Tang, Song-Chun Zhu, Yitao Liang, and Muhan Zhang. 2024. Rule: Knowledge graph reasoning with rule embedding. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 4316–4335.

Komal Teru, Etienne Denis, and Will Hamilton. 2020. Inductive relation prediction by subgraph reasoning. In *International conference on machine learning*, pages 9448–9457. PMLR.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR.

Liang Wang, Wei Zhao, Zhuoyu Wei, and Jingming Liu. 2022. Simkgc: Simple contrastive knowledge graph completion with pre-trained language models. *arXiv preprint arXiv:2203.02167*.

Yu Wang, Nedim Lipka, Ryan A Rossi, Alexa Siu, Ruiyi Zhang, and Tyler Derr. 2024. Knowledge graph prompting for multi-document question answering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 19206–19214.

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.

Yanbin Wei, Qiushi Huang, James T Kwok, and Yu Zhang. 2024. Kicgpt: Large language model with knowledge in context for knowledge graph completion. *arXiv preprint arXiv:2402.02389*.

Shelvia Wongso, Rohan Ghosh, and Mehul Motani. 2023. Using sliced mutual information to study memorization and generalization in deep neural networks. In *AISTATS*, pages 11608–11629.

Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. 2016. Representation learning of knowledge graphs with entity descriptions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

Bo Xue, Yi Xu, Yunchong Song, Yiming Pang, Yuyang Ren, Jiaxin Ding, Luoyi Fu, and Xinbing Wang. 2024. Unlock the power of frozen llms in knowledge graph completion. *arXiv e-prints*, pages arXiv–2408.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Kg-bert: Bert for knowledge graph completion. *arXiv preprint arXiv:1909.03193*.

Liang Yao, Jiazhen Peng, Chengsheng Mao, and Yuan Luo. 2025. Exploring large language models for knowledge graph completion. In *ICASSP*, pages 1–5. IEEE.

Xiao Zhang, Miao Li, and Ji Wu. 2024a. Co-occurrence is not factual association in language models. *Advances in Neural Information Processing Systems*, 37:64889–64914.

Yichi Zhang, Zhuo Chen, Lingbing Guo, Yajing Xu, Wen Zhang, and Huajun Chen. 2024b. Making large language models perform better in knowledge graph completion. In *ACM MM*, pages 233–242.

Andy Zou, Long Phan, Sarah Chen, James Campbell, Phillip Guo, Richard Ren, Alexander Pan, Xuwang Yin, Mantas Mazeika, Ann-Kathrin Dombrowski, and 1 others. 2023. Representation engineering: A top-down approach to ai transparency. *arXiv preprint arXiv:2310.01405*.

A Related Work

Knowledge graph reasoning aims to infer missing triples from incomplete KGs (Tang et al., 2024). Classical methods learn embedding-based scoring functions over graph topology, including translational models (e.g., TransE (Bordes et al., 2013), RotatE (Sun et al., 2019)), semantic matching models (e.g., DistMult (Yang et al., 2014), ComplEx (Trouillon et al., 2016), TuckER (Balažević et al., 2019)), and convolutional models (e.g., ConvE (Dettmers et al., 2018)). While effective at capturing local structural patterns, these methods struggle with data sparsity.

Language model-based methods address these limitations by incorporating textual semantics. Early encoder-only approaches, including KG-BERT (Yao et al., 2019) and LASS (Shen et al., 2022a), reformulate KGC as sequence classification. Later work enhances this paradigm through multi-task learning (Kim et al., 2020) and contrastive objectives (Wang et al., 2022). Generative methods such as KG-S2S (Chen et al., 2022a) and KGT5 (Saxena et al., 2022) adopt generative paradigms with encoder-decoder architectures. CSProm-KG (Chen et al., 2023) introduces conditional soft prompts to balance structural and textual knowledge. Recent advances leverage decoder-only LLMs for KGC. KoPA (Zhang et al., 2024b) combines structural embeddings with Alpaca for triple classification, KG-LLAMA (Yao et al., 2025) frames KGR as instruction-following question answering with parameter-efficient fine-tuning and FLAME leverages frozen LLM representations with task-specific KGC classifiers.

However, existing LM-based methods primarily rely on supervised fine-tuning over data with explicit entity-relation co-occurrences, which biases models toward memorizing surface-level statistical patterns rather than relational reasoning. In contrast, RADAR co-designs the task formulation, training objective, and inference mechanism to prioritize generalization over surface-form memorization in KGR.

B Datasets and Metrics Details

We evaluate the performance of RADAR on four benchmark datasets, covering two distinct tasks: link prediction and triple classification. The statistics of these datasets are summarized in Table 7. For link prediction, we utilize FB15K-237, a subset of Freebase with inverse relations removed to prevent data leakage, and FB15K-237N, a modified version that removes mediator nodes and introduces hard negatives to increase difficulty. For triple classification, we employ WN18RR, a subset of WordNet derived from WN18 by removing inverse relations; UMLS, a medical semantic network consisting of biomedical concepts and their semantic relations; and the aforementioned FB15K-237N.

Table 7: Dataset statistics.

Dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	**# Train**	**# Valid**	**# Test**
FB15K-237	14,541	237	272,115	17,535	20,466
FB15K-237N	13,104	93	87,282	7,041	8,226
WN18RR	40,943	11	86,835	6,068	6,268
UMLS	135	46	5,216	1,304	1,322

For link prediction, we report the Mean Reciprocal Rank (MRR) and Hits@k ($k = 1, 3, 10$) following standard protocols. MRR is the average of the reciprocal ranks of the correct entities, while Hits@k measures the proportion of correct entities ranked in the top-k. For triple classification, we report accuracy. To ensure a fair evaluation, the test sets for triple classification are constructed with an equal number of positive and negative triples.

C Additional Implementation Details

For each query $(h, r, ?)$, we construct a discriminative candidate set $\mathcal{C}(h, r)$ with a fixed size of $K = 4$, containing both positive and negative tail entities as described in Section 2.2.

Entity descriptions are sourced from Xie et al. (2016) for FB15K-237 and FB15K-237N, synset definitions from Yao et al. (2019) for WN18RR, and the original dataset for UMLS. Both Stage I SFT and Stage II GRPO are implemented using LoRA for parameter-efficient fine-tuning, with all backbone parameters kept frozen. For representation-based inference, we consistently extract hidden states from an intermediate layer with index $l = 15$, following prior work showing that intermediate representations better preserve relational knowledge.

For link prediction, exhaustively scoring the entire entity set $\mathcal{E}$ is computationally prohibitive for LLMs. We therefore adopt a retrieve-then-rerank strategy (Li et al., 2025; Wei et al., 2024), utilizing a lightweight pre-trained TuckER model to retrieve the top-$n = 15$ candidates, which are subsequently reranked by our learned classifier. For MRR calculation, should the ground truth entity fall outside

this retrieved top-n set, we use the rank provided by the retriever. While this imposes a recall bottleneck, it balances large-scale inference efficiency with precision, offering a more rigorous evaluation than ranking against randomly sampled negatives (Yao et al., 2025). Crucially, this reliance on retrieval is exclusive to link prediction task. Our triple classification results (Table 2) and the core analytical experiments in Section 3 are conducted without retrieval, directly validating the intrinsic discriminative reasoning capabilities of RADAR.

Table 8: Link classification performance on FB15K-237N under varying parameter n.

Parameter n	MRR
15	**0.415**
20	0.402
25	0.393

D Additional Details on Task-Adaptive Information Quantification

Section 2.5 introduces a task-adaptive mutual information metric for quantifying how much KGR-relevant signal is encoded in intermediate representations. Here we provide additional motivation for this design and explain why standard SMI may be suboptimal in our setting.

Standard SMI approximates $I(X;Y)$ by averaging mutual information over random one-dimensional projections. However, in LLM representations, task-relevant information often concentrates in low-dimensional subspaces (Hu et al., 2022), and random projections dilute this signal by uniformly sampling directions regardless of their informativeness. This leads to noisier estimates with weaker task alignment. Moreover, restricting to linear projections can be suboptimal when the relevant structure is expressed through nonlinear feature interactions.

To address these limitations, We leverage the trained probe f_ϕ (Section 2.4) to define task-aligned projections. Given N samples with representations $\mathbf{Z}^{(l)} \in \mathbb{R}^{d\times N}$, we compute the probe's first-layer activations:

$$\mathbf{V} = \text{PReLU}(\mathbf{W}_1\mathbf{Z}^{(l)} + \mathbf{b}_1) \in \mathbb{R}^{d_v\times N}, \quad (6)$$

where $\mathbf{W}_1 \in \mathbb{R}^{d_v\times d}$ is the learned weight matrix, $\mathbf{b}_1 \in \mathbb{R}^{d_v}$ is the bias vector (with broadcasting across samples), and $\mathbf{V} \in \mathbb{R}^{d_v\times N}$ is the matrix of transformed hidden representations. Crucially, since $\mathbf{W}_1$ is optimized via supervised learning to minimize classification loss on KGR, each row of $\mathbf{W}_1$ defines a projection direction that is implicitly optimized for label discriminability.

This learned transformation differs from random projections in two respects. First, the PReLU nonlinearity enables modeling of nonlinear feature interactions inherent to contextualized embeddings. Second, supervised optimization shapes $\mathbf{W}_1$ to emphasize label-informative subspaces while suppressing task-irrelevant dimensions. As a result, the hidden layer of f_ϕ defines a compact, task-aligned projection space that preserves KGR-relevant information, leading to more focused and better aligned mutual information estimates than random projections.

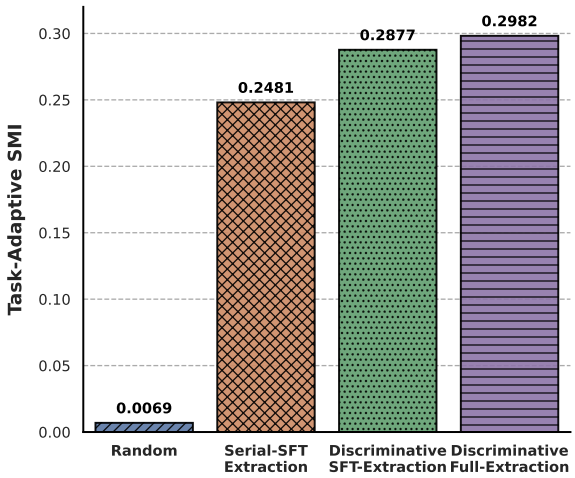

Figure 7: Task-Adaptive SMI of intermediate representations across different methods on FB15K-237N.

E Additional Details on Training

Stage I: Supervised Fine-Tuning. In Stage I, we perform supervised fine-tuning using structured CoT rationales to guide relational reasoning. For each entity, we use a long-form textual description $d(\cdot)$. Prompts are constructed using a fixed template that presents the semantics of the head entity and all candidate entities, followed by the answer selection instruction. The prompts are shown in Figure 8:

Formally, given a training dataset $\mathcal{D}_{\text{train}}$, where each instance consists of a discriminative prompt x constructed from $(h, r, \mathcal{C}(h, r))$ and a target sequence y comprising the CoT reasoning and the final answer. We optimize the model parameters θ using the standard next-token prediction objective:

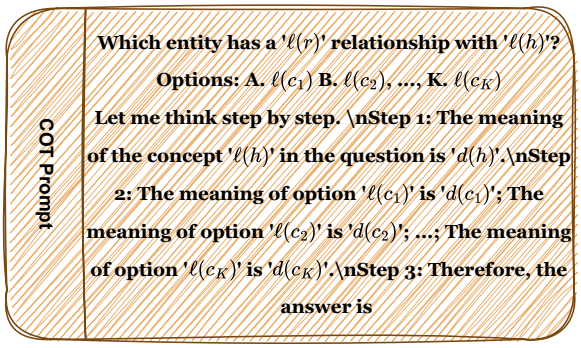

Figure 8: COT prompt.

$$\mathcal{L}_{\text{SFT}} = -\mathbb{E}_{(x,y)\sim\mathcal{D}_{\text{train}}}\left[\sum_{t=1}^{|y|}\log p_\theta(y_t \mid x, y_{<t})\right] \tag{7}$$

Stage II: Reinforcement Learning. We optimize the policy using GRPO, which updates the model by comparing the relative quality of sampled responses for each input. For each query, the model samples a group of G candidate responses $\{\hat{y}_i\}_{i=1}^{G}$, each scored by $R(x, \hat{y}_i)$. Rewards are normalized within each group to compute advantages that guide policy updates. The objective is:

$$\begin{aligned}\mathcal{J}_{\text{GRPO}}(\theta) =& \mathbb{E}_{x\sim\mathcal{D}_{\text{error}},\{\hat{y}_i\}\sim\pi_{\theta_{\text{old}}}}\left[\frac{1}{G}\sum_{i=1}^{G}L_i^{clip}\right.\\ &\left.-\beta\,\mathbb{D}_{\text{KL}}(\pi_\theta \,||\, \pi_{\text{ref}})\right],\end{aligned} \tag{8}$$

$$L_i^{clip} = \min\left(w_i A_i, \text{clip}(w_i, 1-\epsilon, 1+\epsilon)A_i\right), \tag{9}$$

where $w_i = \frac{\pi_\theta(\hat{y}_i|x)}{\pi_{\theta_{\text{old}}}(\hat{y}_i|x)}$, $\pi_{\theta_{\text{old}}}$ is the policy before the update, π_{ref} is the reference policy (the initial SFT model), ϵ and β are hyperparameters controlling update clipping and KL regularization, and A_i denotes the normalized advantage of each candidate within its group based on the reward $R(x, \hat{y}_i)$.

F Additional Ablation Results

Figure 9 shows that as the layer depth increases, the effectiveness of hidden states for triple classification exhibits a unimodal trend, with performance first improving and then degrading. Hidden states from intermediate and upper layers generally yield higher prediction accuracy than those from lower layers. This observation is consistent with prior findings (Geva et al., 2020), which show

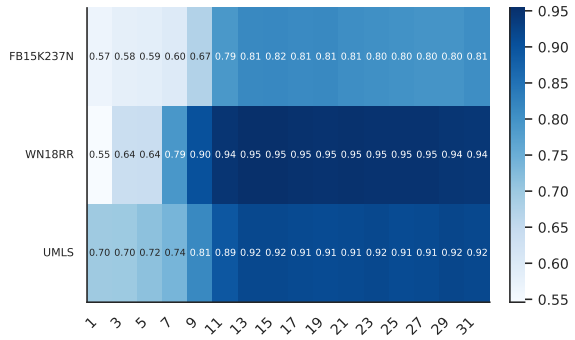

Figure 9: Layer-wise triple classification accuracy of RADAR across datasets. The horizontal axis denotes the layer depth in LLaMA, and the vertical axis corresponds to different datasets.

Table 9: Triple classification accuracy on UMLS under varying parameter K of Discriminative-SFT-Extraction. The best result is highlighted in bold.

Parameter K	Accuracy
3	0.892
4	**0.898**
5	0.887
6	0.893

that Transformer-based language models encode knowledge hierarchically across layers, with higher layers progressively integrating information from lower-level representations. During pretraining, the lower layers may not have stored certain knowledge, thus failing to produce hidden states informative for triple classification. Furthermore, layers closest to the output head tend to underperform relative to middle layers in triple classification accuracy. This phenomenon may be attributed to the capacity of intermediate layers to strike an optimal balance between semantic abstraction and information retention. While lower layers contain raw input noise and upper layers converge toward generic token-prediction objectives, intermediate layers distill task-salient features that are most conducive to discriminative relational reasoning (Jiang et al., 2024; Zou et al., 2023).

Finally, we investigate the impact of the candidate set size K in Table 9. While increasing K raises task difficulty—potentially necessitating more robust representations to distinguish the ground truth from a larger pool of distractors—we observe that performance peaks at $K = 4$ and subsequently plateaus. This phenomenon parallels our findings on negative hardness in Table 3, where escalating difficulty from Tier 2 (Medium) to Tier

3 (Hard) yields no significant performance gain.

These results collectively suggest that the primary driver of our method's effectiveness is not the severity of the classification task, but the mechanism of discrimination itself. The critical factor is the paradigm shift: by constraining the LLM to select among discrete options, we successfully sever the reliance on generative co-occurrence shortcuts and force the activation of relational reasoning circuits. Once this discriminative mode is engaged, the model learns to isolate relational semantics; further increasing the complexity of the candidate space (via larger K or harder negatives) offers diminishing returns and may introduce optimization noise rather than stronger learning signals.