

Beyond Passive Debate: Intent-Driven Multi-Agent Deliberation for Reasoning under Incomplete Knowledge Graphs

Yongqiang Liu¹, Qiyao Peng^{2*}, Binrong Liu¹, Jiawei Chen¹, Xuewei Li¹, Ning Yuan³

¹College of Intelligence and Computing, Tianjin University, Tianjin, China

²School of Artificial Intelligence, Tianjin University, Tianjin, China

³School of Intelligent Computing Engineering, Tianjin Ren'ai College

{lyq236, qypeng, binrong, ppw, lixuewei, ningyuan}@tju.edu.cn

Abstract

Large Language Model (LLM) agents have established a promising paradigm for Knowledge Graph Question Answering (KGQA). However, their practical performance is severely constrained by the pervasive incompleteness of real-world knowledge graphs (KGs). To address this, Incomplete KGQA (IKGQA) has emerged as a vital task, yet existing multi-agent frameworks often suffer from a logic-misalignment effect, where passive collaboration leads to retrieval failure and hallucination propagation when encountering structural gaps. In this paper, we propose **IDLC**, an **Intent-Driven Logical Compensation** framework designed for faithful IKGQA. Unlike traditional reactive debate, IDLC formalizes reasoning as a proactive deliberation task within a Dec-POMDP. It introduces an Active KG Explorer that generates explicit Gap Descriptors to translate structural gaps into actionable retrieval goals, alongside a Logical Consistency Verifier that performs bi-directional anchoring to intercept step-wise semantic drift. The entire multi-agent policy is jointly optimized via reinforcement learning with a specialized logic-fidelity reward, aligning agent behaviors toward faithful reasoning paths. Extensive experiments on BRINK and IKG-CWQ/WebQSP benchmarks demonstrate that IDLC outperforms state-of-the-art baselines in both reasoning performance and step-wise logical fidelity. Furthermore, IDLC exhibits robust zero-shot transferability across various incompleteness levels while maintaining computational efficiency, confirming the effectiveness of intent-driven coordination in repairing fractured reasoning chains.

1 Introduction

The integration of Large Language Models (LLMs) and structured Knowledge Graphs (KGs) establishes a promising paradigm for Knowledge Graph Question Answering (KGQA) (Chen et al., 2024;

Sun et al., 2024; Tan et al., 2025; Shen et al., 2026). By executing multi-step inference over task-pertinent subgraphs, these frameworks leverage relational precision to steer generative processes (LUO et al., 2024; Jiang et al., 2025). However, practical performance remains constrained by pervasive KG incompleteness, where critical factual links are frequently missing due to rapid information evolution and high maintenance costs (Xue and Zou, 2022; Xu et al., 2024; Li et al., 2025). To sustain reasoning integrity under these constraints, Incomplete KGQA (IKGQA) has emerged to bridge structural gaps with external knowledge (Xu et al., 2024; Liu et al., 2026a).

Existing IKGQA methods typically augment context with unstructured corpora or LLM’s internal parametric knowledge (Sun et al., 2019; Lv et al., 2020; Xu et al., 2024; Zhou et al., 2026). While recent multi-agent paradigms (Tran et al., 2025; Gu et al., 2025) introduce specialized KG and RAG agents, they often suffer from “retrieval failure” due to reactive, parallel execution rather than proactive synergy (Chen et al., 2025b; Zhou et al., 2026). When encountering structural gaps, KG agents fail to provide explicit guidance, causing a disconnect where retrieved evidence cannot bridge these specific gaps (Zhou et al., 2026). This **passive collaboration** induces a **logic-misalignment effect**, unanchoring retrieval from structural deficiencies and cascading errors in multi-hop reasoning.

As established in recent Multi-Agent Reinforcement Learning (MARL) studies (Chen et al., 2025b; Liao et al., 2025; Wang et al., 2025b), optimizing isolated components often misaligns individual agent policies with the global objective of logical soundness (Park et al., 2025; Yang and Thomason, 2026). In IKGQA, this manifests as a failure to recover latent alternative reasoning trajectories (Zhou et al., 2026). Consequently, we transition from passive debate to intent-driven collaboration optimized

*Corresponding Author.

via joint reinforcement learning, motivating every retrieval by specific structural gaps.

To this end, we propose **IDLC**¹, an **Intent-Driven Logical Compensation** framework for faithful IKGQA. Structurally distinct from traditional reactive debate, IDLC formalizes reasoning as a proactive deliberation task within a Decentralized Partially Observable Markov Decision Process (Dec-POMDP). It introduces an *Active KG Explorer* generating explicit *Gap Descriptors*—logical intents translating structural gaps into actionable retrieval goals. To sustain integrity, a *Logical Consistency Verifier* performs bi-directional anchoring between symbolic paths and semantic evidence, triggering a corrective state to prune fallacious trajectories. The pipeline is jointly optimized via Multi-Agent PPO (MAPPO) (Yu et al., 2022) with a specialized *logic-fidelity reward* to align agent behaviors toward faithful alternative paths. In summary, our main contributions are:

- We formalize the logic-misalignment effect in passive multi-agent collaboration and propose IDLC, a novel MARL framework transitioning from passive debate to intent-driven logical compensation, utilizing explicit Gap Descriptors to proactively repair fractured reasoning chains.
- We introduce a bi-directional Logical Consistency Verifier optimized via a specialized logic-fidelity reward mechanism, which effectively coordinates agent behaviors within a Dec-POMDP to prioritize long-term consistency and sustain high faithfulness across multi-hop reasoning.
- Extensive experiments on BRINK and IKG-CWQ/WebQSP benchmarks demonstrate that IDLC achieves state-of-the-art performance and step-wise stability in suppressing long-range error accumulation. Furthermore, it exhibits robust zero-shot transferability across various incompleteness levels while reducing token overhead via adaptive retrieval.

2 Related Work

LLM-based IKGQA. To enhance generative faithfulness, recent research integrates LLMs with structured KGs (Sun et al., 2024; Chen et al., 2024; Tan et al., 2025; Shen et al., 2026), which are classified into: (i) *knowledge-internalization*, em-

bedding KG structures via fine-tuning (Luo et al., 2024, 2025); and (ii) *knowledge-interaction*, where LLMs dynamically query external symbolic structures (Sun et al., 2024; Jiang et al., 2025). Though effective, they remain hampered by pervasive KG incompleteness (Xue and Zou, 2022; Xu et al., 2024; Liu et al., 2026a). Current IKGQA methods bridge structural gaps via external corpora augmentation (Sun et al., 2019; Xu et al., 2024) or adaptive fusion like DoM (Liu et al., 2026a). Distinct from these approaches neglecting latent structural priors, IDLC leverages explicit Gap Descriptors to identify logical fractures, sustaining reasoning integrity.

Multi-Agent Collaboration. The multi-agent paradigm enhances reliability by coordinating specialized agents via consensus-driven or deliberative processes (Du et al., 2024; Liang et al., 2024; Yang and Thomason, 2026). Recent advances transition from rigid, fixed-turn protocols toward inference-adaptive self-organizing systems (Hu et al., 2025; Zhang et al., 2025b,a). However, most frameworks treat agents as passive executors of predefined schemes, incurring a “Meta-Cognitive Blindspot” where they fail to adjust strategies based on internal cognitive states (Yang and Thomason, 2026). In IKGQA, this manifests as passive collaboration, where KG and RAG agents operate in parallel without proactive alignment. IDLC resolves this deficiency via an intent-driven mechanism orchestrating proactive logical compensation between agents.

Multi-Agent Reinforcement Learning. Research increasingly adopts MARL to optimize agent coordination beyond heuristic protocols (Chen et al., 2025b; Liu et al., 2026b; Yang and Thomason, 2026). Frameworks like MAPoRL (Park et al., 2025) and LERO (Wei et al., 2025) employ collaborative rewards for complex reasoning co-training. To mitigate reward instability, Soft-RankPO (Yang and Thomason, 2026) introduces rank-based advantage shaping, whereas MAGRPO (Liu et al., 2026b) leverages group-relative policy optimization (Shao et al., 2024; Chen et al., 2025a) within a Dec-POMDP framework (Oliehoek et al., 2016) to stabilize joint policies. Furthermore, MMOA-RAG (Chen et al., 2025b) proves that joint optimization mitigates retrieval-generation goal misalignment. Building upon these, IDLC applies MARL to align agent actions with structural topology, optimizing logical compensation and step-wise consistency in incomplete environments.

¹Our source code and datasets are available at <https://anonymous.4open.science/r/IDLC-W4E7>

3 Preliminary

IKGQA Task. Standard KGQA aims to predict correct answer entities $A_q \subseteq E$ for a natural language question q over a complete KG $G = (E, R, T)$ (Sun et al., 2024). However, real-world KGs are frequently incomplete due to construction and maintenance limitations, causing fragmented or non-existent reasoning paths (Liu et al., 2026a; Zhou et al., 2026). To mitigate this, IKGQA relaxes full KG coverage constraints by enabling the reasoning process to leverage external knowledge $\mathcal{D}$ (e.g., textual corpora or LLM internal parametric knowledge) (Xu et al., 2024; Liu et al., 2026a). Formally, the IKGQA task learns a mapping $f : (q, G, \mathcal{D}) \mapsto A_q$ that sustains reasoning integrity by bridging structural gaps in G with semantic evidence from $\mathcal{D}$.

Multi-Agent Reasoning as a Dec-POMDP. To facilitate proactive coordination and joint optimization, we formalize the IKGQA reasoning process as a **Decentralized Partially Observable Markov Decision Process (Dec-POMDP)**, defined by the tuple $(\mathcal{I}, \mathcal{S}, \mathcal{A}, \mathcal{T}, R, \Omega, \mathcal{O}, \gamma)$. Specifically, $\mathcal{I}$ denotes heterogeneous agents acting within global state space $\mathcal{S}$ and joint action space $\mathcal{A}$. $\mathcal{T} : \mathcal{S} \times \mathcal{A} \mapsto \mathcal{S}$ and $R : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ represent state transition and shared reward functions, respectively. Ω and $\mathcal{O} : \mathcal{S} \times \mathcal{I} \mapsto \Omega$ characterize observation space and the function providing each agent with a partial, local view of the reasoning context; $\gamma \in [0, 1]$ is the discount factor. This formulation transforms multi-hop reasoning from a heuristic-driven static protocol into a learned, dynamic deliberation strategy optimized for long-term consistency.

4 Methodology

In this section, we propose **IDLC**, an **Intent-Driven Logical Compensation** framework based on multi-agent reinforcement learning for faithful IKGQA. As illustrated in Figure 1, IDLC conceptualizes the reasoning process as a collaborative deliberation task within a Dec-POMDP $(\mathcal{I}, \mathcal{S}, \mathcal{A}, \mathcal{T}, R, \Omega, \mathcal{O}, \gamma)$, orchestrating a “Grounding-Deliberation-Verification” loop. Specifically, the grounding mechanism first identifies candidate entities; building upon this, three specialized agents $i \in \mathcal{I}$ collaboratively navigate the graph, compensate for missing logic, and verify reasoning faithfulness, namely: (i) an Active KG Ex-

plorer ($a_{x,t} \in \mathcal{A}_x$), (ii) a RAG Agent ($a_{r,t} \in \mathcal{A}_r$), and (iii) a Logical Consistency Verifier ($a_{v,t} \in \mathcal{A}_v$). The entire system is jointly optimized for long-term consistency and faithfulness, while initializing an empty predicted answer set $\mathcal{Y}_{ans} = \emptyset$ for final answer generation.

4.1 Question Anchor Grounding

To initiate reasoning, IDLC grounds the question q by identifying a candidate anchor set $\mathcal{E}_{anc} \subset E$. Following the semantic grounding strategy (Xu et al., 2026), we encode all KG entity names using MiniLM (Wang et al., 2020) and construct a FAISS index (Douze et al., 2025) for efficient similarity search. Specifically, keyword mentions W are extracted directly from q to retrieve the initial candidate set $\tilde{E}$ via the constructed index.

To disambiguate these candidates, we evaluate the semantic alignment between the local neighborhood $\mathcal{N}(e)$ of each $e \in \tilde{E}$ and the question q . Using a pre-trained SBERT model (Reimers and Gurevych, 2019), we compute a relevance score $S_{anc}(e)$ as:

$$S_{anc}(e) = \sum_{r \in \mathcal{N}(e)}^{top-k} \cos(\mathbf{q}, \mathbf{r}), \quad (1)$$

where $\mathbf{q}$ and $\mathbf{r}$ are SBERT embeddings of the question and relation, respectively. The top- m entities with the highest $S_{anc}(e)$ form the final anchor set $\mathcal{E}_{anc}$.

4.2 Intent-Driven Deliberation and Compensation

The deliberation and compensation phase involves synergistic symbolic exploration and semantic compensation. To handle multiple entry points, IDLC initiates parallel exploration by assigning each entity $e \in \mathcal{E}_{anc}$ as the starting node of an independent reasoning trajectory, unfolding as follows:

Proactive Exploration on KG. Let e_t denote the current entity at step t within a trajectory, initialized with $e_1 = e$ (where $e \in \mathcal{E}_{anc}$). The Active KG Explorer probes the global state $s_t \in \mathcal{S}$ through local observations $o_{x,t} = \mathcal{O}(s_t, x) \in \Omega$, where $x \in \mathcal{I}$ denotes the Active KG Explorer, to determine if the logic chain can be extended via symbolic relations. We define a reasoning context $\mathcal{C}_t = \{q, p_t\}$ as the union of the question and current reasoning path p_t . To identify structural gaps, the Gap Indicator

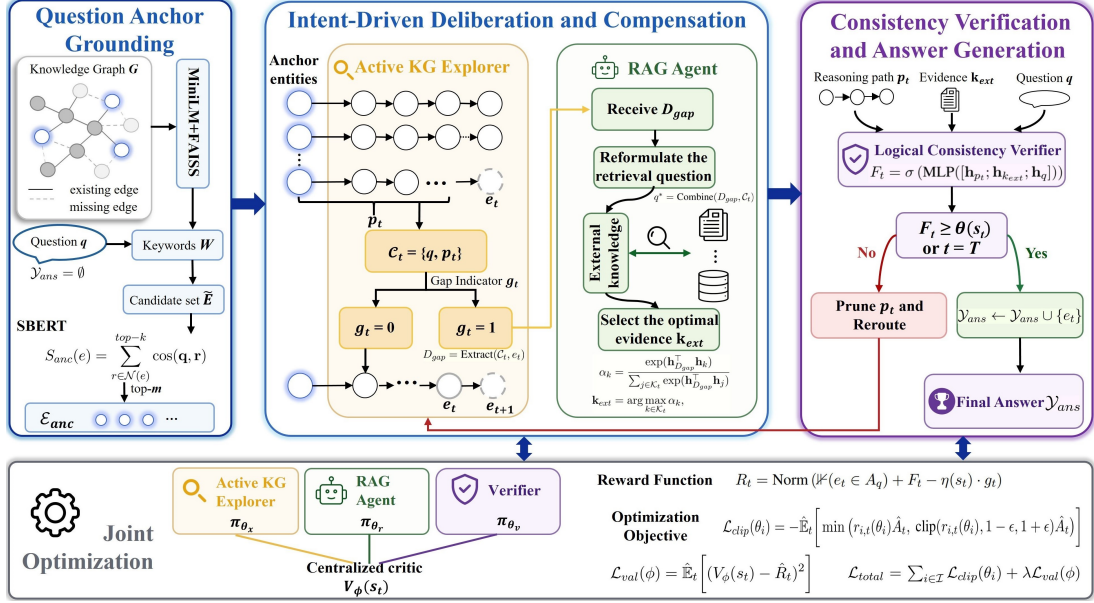


Figure 1: The overall framework of IDLC.

$g_t \in \{0, 1\}$ is defined as:

$$g_t = \mathbb{I} \left(\max_{r \in \mathcal{N}(e_t)} \text{Sim}(\phi(r), \psi(C_t)) < \tau(s_t) \right), \quad (2)$$

where $\phi(r)$ denotes the embedding of relation $r \in \mathcal{N}(e_t)$, and $\psi(C_t)$ denotes the semantic intent representation parameterized by the Explorer’s policy network based on $o_{x,t}$. To adapt to varying graph densities, we utilize an adaptive threshold $\tau(s_t) = \sigma(W_\tau s_t + b_\tau)$, where s_t is the state vector, $\sigma(\cdot)$ is the sigmoid activation, and W_τ, b_τ are learnable parameters. When $g_t = 0$, the Explorer samples a symbolic relation action $a_{x,t} \in \mathcal{A}_x$ to extend the reasoning path. Conversely, when $g_t = 1$, the Explorer generates a Gap Descriptor $D_{gap} = \text{Extract}(C_t, e_t)$ representing the missing logical intent.

Logic Compensation via RAG. Upon receiving D_{gap} , the RAG Agent performs targeted retrieval from external knowledge $\mathcal{D}$, reformulating the retrieval question as $q^* = \text{Combine}(D_{gap}, C_t)$. The agent selects the optimal evidence $\mathbf{k}_{ext}$ from candidate snippets $\mathcal{K}_t$ by maximizing the alignment score α_k :

$$\alpha_k = \frac{\exp(\mathbf{h}_{D_{gap}}^\top \mathbf{h}_k)}{\sum_{j \in \mathcal{K}_t} \exp(\mathbf{h}_{D_{gap}}^\top \mathbf{h}_j)}, \quad (3)$$

$$\mathbf{k}_{ext} = \arg \max_{k \in \mathcal{K}_t} \alpha_k,$$

where $\mathbf{h}_{D_{gap}}^\top$ and $\mathbf{h}_k$ are latent embeddings of the descriptor and evidence, respectively.

4.3 Consistency Verification and Answer Generation

The Logical Consistency Verifier evaluates step-wise faithfulness by cross-validating the proposed reasoning path p_t against retrieved evidence $\mathbf{k}_{ext}$, computing a Fidelity Score F_t through a feature fusion module:

$$F_t = \sigma(\text{MLP}([\mathbf{h}_{p_t}; \mathbf{h}_{\mathbf{k}_{ext}}; \mathbf{h}_q])), \quad (4)$$

where $[\cdot; \cdot]$ denotes vector concatenation. IDLC employs an adaptive verification threshold $\theta(s_t) = \sigma(W_\theta s_t + b_\theta)$. The step-wise routing and final answer set ($\mathcal{Y}_{ans}$) generation are formally defined as:

$$\begin{cases} \text{Prune } p_t \text{ and Reroute,} & \text{if } F_t < \theta(s_t) \\ \mathcal{Y}_{ans} \leftarrow \mathcal{Y}_{ans} \cup \{e_t\}, & \text{if } F_t \geq \theta(s_t) \vee t = T. \end{cases} \quad (5)$$

Once the terminal condition is triggered, the validated reasoning paths and external evidence $\mathbf{k}_{ext}$ are compiled to generate the final answer.

4.4 Joint Optimization

To optimize strategic interactions within the Dec-POMDP, we employ the MAPPO framework (Yu et al., 2022; Yang and Thomason, 2026), where each agent $i \in \mathcal{I}$ maintains a policy π_{θ_i} and shares a centralized critic $V_\phi(s_t)$.

Reward Function. The shared reward R_t aligns agents toward resolving structural gaps while maintaining efficiency:

$$R_t = \text{Norm}(\mathbb{I}(e_t \in A_q) + F_t - \eta(s_t) \cdot g_t), \quad (6)$$

where $\eta(s_t) = \text{softplus}(W_\eta s_t + b_\eta)$ is a learned penalty coefficient with learnable parameters W_η and b_η , and A_q is the ground-truth answer set. $\text{Norm}(\cdot)$ stabilizes the multi-agent training signals via moving average normalization.

Optimization Objective. Each agent i updates its policy by minimizing the clipped surrogate loss $\mathcal{L}_{clip}(\theta_i)$:

$$\mathcal{L}_{clip}(\theta_i) = -\hat{\mathbb{E}}_t \left[\min \left(r_{i,t}(\theta_i) \hat{A}_t, \right. \right. \\ \left. \left. \text{clip}(r_{i,t}(\theta_i), 1 - \epsilon, 1 + \epsilon) \hat{A}_t \right) \right], \quad (7)$$

where $r_{i,t}(\theta_i) = \pi_{\theta_i}(a_{i,t}|o_{i,t})/\pi_{\theta_i^{\text{old}}}(a_{i,t}|o_{i,t})$ is the importance sampling ratio, $a_{i,t} \in \mathcal{A}_i$ and $o_{i,t} = \mathcal{O}(s_t, i) \in \Omega$ denote the local action and observation of agent $i \in \{x, r, v\}$ at step t respectively, $\hat{A}_t$ is the advantage function, and ϵ is the clipping hyper-parameter. Simultaneously, the centralized critic V_ϕ is updated by minimizing the value loss:

$$\mathcal{L}_{val}(\phi) = \hat{\mathbb{E}}_t \left[(V_\phi(s_t) - G_t)^2 \right], \quad (8)$$

where $G_t = \sum_{l=0}^{T-t} \gamma^l R_{t+l}$ is the discounted return representing the cumulative future rewards from the global state s_t , and γ is the discount factor. The total objective $\mathcal{L}_{total} = \sum_{i \in \mathcal{I}} \mathcal{L}_{clip}(\theta_i) + \lambda \mathcal{L}_{val}(\phi)$ ensures policy convergence, where λ is a weight coefficient. The complete execution logic of IDLC is formally detailed in Algorithm 1 (Appendix A).

5 Experiments

5.1 Experimental Setup

Datasets and Evaluation. We evaluate IDLC on two benchmark categories: (1) **BRINK** (Zhou et al., 2026), comprising **Family-Incomplete**, **FB15k-237-Incomplete**, and **Wikidata5m-Incomplete** with Horn-rule grounded incompleteness, utilizing official Train/Val/Test splits for policy optimization and within-domain evaluation. (2) **IKG-CWQ/WebQSP** (Xu et al., 2024), featuring four randomized triple deletion degrees (20%–80%) based on original benchmarks. These datasets assess the cross-dataset robustness of IDLC via zero-shot transfer of policies trained on Wikidata5m-Incomplete (the most complex BRINK dataset). We adopt **Hits@1 (Exact Match)** and **F1-score** as primary metrics. More details are provided in Appendix B.

Baselines. We compare IDLC against a representative suite of baselines categorized into three technical paradigms: (1) *Fine-tuned Methods*, including RoG (LUO et al., 2024), ChatKBQA (Luo et al., 2024), and GNN-RAG (Mavromatis and Karypis, 2025); (2) *Symbolic Exploration Frameworks*, such as ToG (Sun et al., 2024), PoG (Tan et al., 2025), and ReKnoS (Wang et al., 2025a); and (3) *Collaborative Reasoning Agents*, featuring GoG (Xu et al., 2024), AnchorRAG (Xu et al., 2026), and DoM (Liu et al., 2026a). Detailed baseline descriptions are provided in Appendix C.

Implementation Detail. We instantiate agents with *Llama-3.1-8B-Instruct*, *Llama-3.2-3B-Instruct* (Grattafiori et al., 2024), *Qwen2.5-7B-Instruct*, and *Qwen2.5-3B-Instruct* (Team, 2025). For KG grounding, we employ SBERT (all-MiniLM-L6-v2) (Reimers and Gurevych, 2019) and FAISS (Douze et al., 2025) for embedding and efficient vector indexing. We set the maximum reasoning depth $T = 4$ to accommodate the multi-hop complexity of the benchmarks. Following Xu et al. (2026), the parallel anchor set size m and top- k relations for semantic scoring (Eq. 1) are fixed at 3 and 5, respectively, with Wikipedia serving as the external knowledge $\mathcal{D}$. All experiments are conducted on four NVIDIA A100 GPUs (80GB), reporting average results over three random seeds. Detailed setups are provided in Appendix D.

5.2 Main Results

As illustrated in Table 1, IDLC consistently achieves state-of-the-art performance across all three IKGQA datasets, outperforming a wide range of baselines. Notably, even with moderate-scale open-source backbones (e.g., 7B/8B parameters), our framework surpasses the strongest baseline, DoM, which utilizes the massive DeepSeek-v3 model. For instance, on the FB15k-237-Incomplete dataset, where baselines generally struggle to exceed 50% Hits@1, IDLC (Qwen2.5-7B) achieves a remarkable 57.5%, demonstrating its superior ability to handle complex relational reasoning under knowledge constraints. This consistent superiority confirms that our intent-driven logic compensation strategy is highly effective at mitigating the structural gaps of incomplete KGs.

Performance with Different Backbones. We evaluate the impact of varied LLM backbones on reasoning performance within the IDLC frame-

Method	Model	Family-Incomplete		FB15k-237-Incomplete		Wikidata5m-Incomplete	
		Hits@1	F1	Hits@1	F1	Hits@1	F1
ChatKBQA	Llama-2-7B	45.1	43.3	28.2	26.5	25.3	24.7
RoG	Llama-2-7B	64.5	57.9	41.5	35.9	38.9	33.5
GNN-RAG	Llama-2-7B	67.3	61.4	43.1	37.5	49.3	42.9
ToG	DeepSeek-v3	61.1	–	41.1	–	42.3	–
ReKnoS	DeepSeek-v3	63.4	–	41.8	–	45.0	–
PoG	DeepSeek-v3	66.6	–	43.9	–	50.7	–
GoG	DeepSeek-v3	70.3	–	47.5	–	51.0	–
AnchorRAG	DeepSeek-v3	71.5	–	48.1	–	50.8	–
DoM	DeepSeek-v3	73.2	–	50.4	–	52.7	–
IDLC	Llama-3.2-3B	71.5	63.8	48.2	41.5	50.5	43.8
	Llama-3.1-8B	75.8	68.4	52.8	46.2	54.8	48.2
	Qwen2.5-3B	76.4	69.1	53.6	47.1	55.6	49.0
	Qwen2.5-7B	80.2	73.4	57.5	51.2	59.2	52.6

Table 1: Performance of IDLC and previous state-of-the-art methods on the three IKGQA datasets.

work. Performance correlates positively with the scale and reasoning capability of the underlying models. While the lightweight Llama-3.2-3B yields lower performance than the DeepSeek-v3-based DoM, it remains highly competitive against exploration-based methods like PoG. Concurrently, Llama-3.1-8B serves as a performance pivot. By optimizing its deliberation policy via MAPPO, this intermediate backbone successfully surpasses the DeepSeek-v3 baseline by an average of 2.4% Hits@1. Furthermore, Qwen2.5-3B outperforms Llama-3.1-8B despite its smaller parameter size, confirming that IDLC effectively leverages the superior instruction-following capabilities of the Qwen architecture to achieve higher reasoning precision.

In-depth Analysis of Graph Incompleteness. The resilience of IDLC is particularly evident when confronting varying degrees of graph incompleteness. We observe that our performance gains over baselines become significantly more pronounced on the highly sparse Wikidata5m-Incomplete compared to the more structured Family-Incomplete. This divergence occurs because traditional symbolic and agentic methods typically suffer from “search paralysis” when encountering structural gaps in incomplete KGs. In contrast, IDLC’s Gap Indicator accurately detects these topological fractures to trigger targeted external compensation from $\mathcal{D}$, successfully repairing broken reasoning chains that would otherwise induce inference stagnation.

Comprehensive Reasoning via F1 Score. To evaluate multi-hop recall and evidence gathering

breadth, we report F1 scores alongside Hits@1. Across all backbone configurations, the F1 score improves in tandem with Hits@1, maintaining a stable and significant margin over previous methods. In IKGQA, achieving a high F1 score is uniquely demanding as it requires retrieving the full spectrum of correct entities despite fractured graph links. The consistently high F1 values (e.g., 73.4% for Qwen2.5-7B on Family-Incomplete) demonstrate that IDLC does not merely exploit single-path reasoning shortcuts. Instead, it effectively deploys logic compensation to recover latent target entities unreachable via internal graph traversal alone.

5.3 Ablation Study

We conduct an ablation study on FB15k-237-Incomplete using Qwen2.5-7B-Instruct to evaluate the contributions of IDLC’s core components (Table 2). The most drastic decline occurs in the *w/o Logic Comp.* variant (47.8% Hits@1), where reasoning chains remain irreparable despite the Gap Indicator still identifying structural gaps (Eq. 2) at a frequency of 1.2 calls per query. This underscores that external compensation from $\mathcal{D}$ is the fundamental prerequisite for bridging structural gaps in incomplete KGs. Notably, the *w/o Gap Indicator* variant, which performs retrieval at every step (2.3 calls), yields a 2.2% lower Hits@1 compared to full IDLC. This confirms that our intent-driven trigger not only reduces computational overhead by almost 50% but also enhances precision by filtering redundant “semantic noise”. Finally, the performance drop and higher retrieval frequency (1.6 calls) of the *w/o MAPPO* variant demonstrate that MARL

is crucial for refining the agent’s deliberation policy. This collaborative optimization enables more decisive and precise calls for assistance than its supervised fine-tuning (SFT) initialized counterpart.

Variant	Hits@1	F1	Avg. Ext. Calls
IDLC	57.5	51.2	1.2
w/o Logic Comp. ($\mathcal{D}$)	47.8	41.2	–
w/o Gap Indicator	55.3	49.1	2.3
w/o MAPPO	53.2	46.8	1.6

Table 2: Ablation study on FB15k-237-Incomplete. Avg. Ext. Calls denotes the average frequency of external knowledge retrieval per query.

5.4 Generalization and Zero-shot Transfer

To evaluate the cross-dataset robustness of IDLC, we perform zero-shot transfer by applying policies trained on Wikidata5m-Incomplete directly to the IKG-40% variants of CWQ and WebQSP. As shown in Table 3, IDLC consistently achieves state-of-the-art performance without task-specific fine-tuning. Crucially, IDLC (Qwen2.5-7B) reaches 65.8% Hits@1 on the complex IKG-CWQ, outperforming the DeepSeek-v3-based DoM by 3.8%. This margin demonstrates that a well-optimized deliberation policy can be more effective than brute-force model parameter scaling for bridging structural gaps in incomplete graphs. While large models like DeepSeek-v3 possess vast parametric knowledge, they remain vulnerable to topological fractures. In contrast, IDLC successfully identifies these gaps to recover reasoning chains through intent-driven compensation. Significant gains in F1 scores (59.2% on IKG-CWQ and 71.5% on IKG-WebQSP) further validate that repairing fractured logic substantially enhances reasoning path recall. This successful transfer confirms that IDLC captures universal semantic cues of reasoning failure rather than overfitting to dataset-specific patterns.

Method	Backbone	IKG-CWQ		IKG-WebQSP	
		Hits@1	F1	Hits@1	F1
GNN-RAG	Llama-2-7B	57.8	51.4	78.5	65.3
PoG	DeepSeek-v3	55.9	–	78.0	–
DoM	DeepSeek-v3	62.0	–	81.7	–
IDLC	Qwen2.5-7B	65.8	59.2	84.3	71.5

Table 3: Zero-shot transfer performance on IKG-CWQ/WebQSP (IKG-40%).

5.5 Efficiency Analysis

To evaluate the practical feasibility of IDLC, we benchmark its computational overhead against

state-of-the-art baselines using 100 randomly sampled queries from the Wikidata5m-Incomplete test set. As illustrated in Table 4, while multi-agent frameworks like DoM achieve robust reasoning, they incur excessive costs (821k tokens and 78 min) due to constant sub-question decomposition and iterative debate. In contrast, IDLC demonstrates superior efficiency, achieving a 34.7% reduction in total tokens compared to DoM. This substantial gain is directly attributed to our intent-driven Gap Indicator, which selectively triggers external compensation only upon detecting topological fractures, eliminating redundant retrieval and verification steps. This proactive gating mechanism maintains a compact context and accelerates the inference cycle (42 min), confirming that IDLC effectively bypasses the conventional accuracy-efficiency trade-off through adaptive resource allocation.

Method	Total Tokens (k)	Total Time (min)
PoG	786	257
GoG	714	63
DoM	821	78
IDLC	536	42

Table 4: Computational cost analysis on 100 samples from Wikidata5m-Incomplete.

5.6 Mechanistic Verification

Evidence & Intent Validity. To dissect IDLC’s internal mechanism, we analyze reasoning success via attribution analysis across four modes: *LLM*, *KG*, *RAG*, and *Hybrid*. As shown in Figure 2, IDLC employs an adaptive integration strategy. On the dense FB15k-237-Incomplete, *Hybrid* reasoning dominates (72.2%), meanwhile *RAG* (19.5%) acts as a structural lifeline on the sparse Wikidata5m-Incomplete. This dynamic shift confirms that our deliberative loop adaptively scales its reasoning depth based on environmental constraints. Qualitative cases (see Appendix E) further reveal that the Explorer generates precise Gap Descriptors to bridge structural gaps. Crucially, the reasoning traces in Table 8 illustrate how the Verifier intercepts low-fidelity candidates ($\mathcal{F}_t < \theta(s_t)$), effectively mitigating hallucination propagation inherent in the local proximity searches of competitive baselines. The high fidelity scores assigned to intent-aligned facts validate the synergy between goal-oriented guidance and evidence-based grounding.

Step-wise Fidelity. To evaluate the long-range reasoning faithfulness of IDLC, we assess the logi-

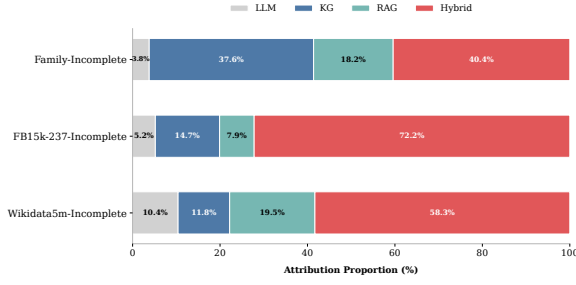


Figure 2: Evidence attribution analysis across three IKGQA datasets.

cal fidelity score ($\mathcal{F}_t$) at each incremental reasoning step. For a rigorous comparison, we utilize our pre-trained Verifier (Eq. 4) to evaluate the intermediate trajectories generated by the competitive baseline, DoM. As illustrated in Figure 3, the verification scores of DoM (blue) rapidly disperse and submerge below the average threshold $\theta(s_t)$ after the second step ($0.93 \rightarrow 0.35$). In contrast, IDLC maintains a steady mean logical fidelity above 0.86 even at step 4. The dense concentration of IDLC’s sample points (red) against the scattered distribution of DoM demonstrates that our corrective mechanism acts as a robust logical filter. By pruning low-confidence branches and re-evaluating alternative symbolic paths, IDLC effectively intercepts error accumulation at its onset, ensuring that the final answer is anchored in a logically coherent evidence chain.

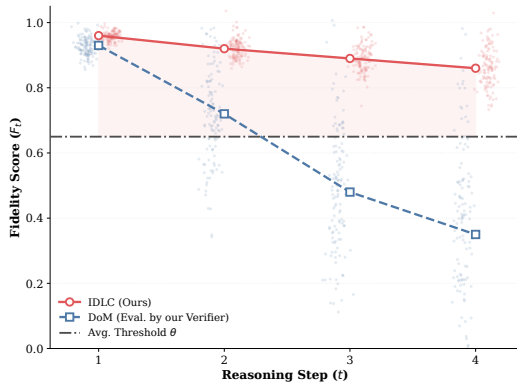


Figure 3: Step-wise fidelity analysis comparing IDLC with DoM under 80% incompleteness level.

5.7 Further Analysis

Impact of Varying Graph Incompleteness. To evaluate IDLC’s generalization under structural degradation, we test its performance in a zero-shot transfer setting across triple deletion rates from 0% to 80%. As observed on IKG-CWQ, the perfor-

mance gap between IDLC and DoM widens from 2.1% at 0% deletion to 4.2% at the extreme 80% incompleteness level. This expanding margin confirms that our intent-driven deliberation effectively mitigates topological fractures via explicit Gap Descriptors. Detailed analyses and robustness curves are provided in Appendix F.

Hyperparameter Sensitivity. To assess the parametric sensitivity of IDLC, we analyze three key parameters governing agent behaviors: candidate evidence pool size $|\mathcal{K}_t|$, maximum reasoning depth T , and discount factor γ . The framework exhibits a consistent trend across both datasets, validating our optimal configuration where performance plateaus at $|\mathcal{K}_t| = 3$, peaks at $T = 4$, and achieves steady gains as γ approaches 0.95. A comprehensive exploration of these parametric behaviors is provided in Appendix G.

Error Analysis. To investigate the performance boundaries of IDLC, we manually categorize 200 failed trajectories from each dataset. The empirical failure distributions are deeply conditioned on the intrinsic topological properties of the KGs. Specifically, sparse structures induce search deadlocks accounting for 42% on Wikidata5m-Incomplete, whereas dense relation sets trigger verification over-pruning accounting for 32% on FB15k-237-Incomplete. The formal taxonomic discussion is presented in Appendix H.

6 Conclusion

In this paper, we present IDLC, a novel intent-driven multi-agent reinforcement learning framework designed to sustain reasoning integrity under incomplete KGs. By transitioning from passive agent debate to a proactive logical compensation paradigm, IDLC effectively bridges structural gaps through explicit Gap Descriptors and adaptive consistency verification for faithful answer generation. Our extensive experiments across multiple IKGQA datasets demonstrate that the synergistic interaction between symbolic exploration and semantic compensation, optimized within a Dec-POMDP, significantly enhances both reasoning performance and step-wise faithfulness. The observed robustness against extreme graph incompleteness and error accumulation confirms that goal-oriented intent alignment is a vital paradigm for building reliable and interpretable reasoning systems in incomplete knowledge environments.

Limitations

We acknowledge that the performance of IDLC is inherently constrained by the coverage of the external knowledge source $\mathcal{D}$, as the logical compensation mechanism may encounter a “knowledge bottleneck” when required facts are absent from $\mathcal{D}$. Furthermore, although IDLC achieves competitive inference efficiency through its adaptive retrieval strategy, the multi-agent deliberation loop and bi-directional verification process inherently possess a higher complexity compared to basic, single-step reasoning protocols. The strategic coordination required for proactive logical repair remains a factor in computational overhead, especially when scaling to extremely long reasoning chains. Future research will focus on further streamlining this deliberation process or distilling the learned collaborative policies into more compact architectures to further minimize the latency-fidelity trade-off.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NO.62602463).

References

- Hongzhan Chen, Tao Yang, Shiping Gao, Ruijun Chen, Xiaojun Quan, Hongtao Tian, and Ting Yao. 2025a. Discriminative policy optimization for token-level reward models. In *International Conference on Machine Learning*, pages 9546–9565. PMLR.
- Liyi Chen, Panrong Tong, Zhongming Jin, Ying Sun, Jieping Ye, and Hui Xiong. 2024. Plan-on-graph: Self-correcting adaptive planning of large language model on knowledge graphs. *Advances in Neural Information Processing Systems*, 37:37665–37691.
- Yiqun Chen, Lingyong Yan, Weiwei Sun, Xinyu Ma, Yi Zhang, Shuaiqiang Wang, Dawei Yin, Yiming Yang, and Jiaxin Mao. 2025b. Improving retrieval-augmented generation through multi-agent reinforcement learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.
- Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2025. The faiss library. *IEEE Transactions on Big Data*.
- Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2024. [Improving factuality and reasoning in language models through multiagent debate](#). In *Forty-first International Conference on Machine Learning*.
- Orri Erling and Ivan Mikhailov. 2009. Rdf support in the virtuoso dbms. In *Networked Knowledge- Networked Media: Integrating Knowledge Management, New Media Technologies and Semantic Systems*, pages 7–24. Springer.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Wenyuan Gu, Jiale Han, Haowen Wang, Xiang Li, and Bo Cheng. 2025. Explain-analyze-generate: A sequential multi-agent collaboration method for complex reasoning. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7127–7140.
- Shengran Hu, Cong Lu, and Jeff Clune. 2025. [Automated design of agentic systems](#). In *The Thirteenth International Conference on Learning Representations*.
- Jinhao Jiang, Kun Zhou, Wayne Xin Zhao, Yang Song, Chen Zhu, Hengshu Zhu, and Ji-Rong Wen. 2025. Kg-agent: An efficient autonomous agent framework for complex reasoning over knowledge graph. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9505–9523.
- Linyu Li, Zhi Jin, Xuan Zhang, Haoran Duan, Jishu Wang, Zhengwei Tao, Haiyan Zhao, and Xiaofeng Zhu. 2025. Multi-view riemannian manifolds fusion enhancement for knowledge graph completion. *IEEE Transactions on Knowledge and Data Engineering*.
- Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2024. Encouraging divergent thinking in large language models through multi-agent debate. In *Proceedings of the 2024 conference on empirical methods in natural language processing*, pages 17889–17904.
- Junwei Liao, Muning Wen, Jun Wang, and Weinan Zhang. 2025. Marft: Multi-agent reinforcement fine-tuning. *arXiv preprint arXiv:2504.16129*.
- Jilong Liu, Pengyang Shao, Wei Qin, Fei Liu, Yonghui Yang, and Richang Hong. 2026a. Debate over mixed-knowledge: A robust multi-agent reasoning framework for incomplete knowledge graph question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 15333–15341.
- Shuo Liu, Zeyu Liang, Xueguang Lyu, and Christopher Amato. 2026b. Llm collaboration with multi-agent reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 32150–32158.
- Haoran Luo, E Haihong, Zichen Tang, Shiyao Peng, Yikai Guo, Wentai Zhang, Chenghao Ma, Guanting

- Dong, Meina Song, Wei Lin, and 1 others. 2024. Chatkbqa: A generate-then-retrieve framework for knowledge base question answering with fine-tuned large language models. In *Findings of the association for computational linguistics: ACL 2024*, pages 2039–2056.
- LINHAO LUO, Yuan-Fang Li, Gholamreza Haffari, and Shirui Pan. 2024. [Reasoning on graphs: Faithful and interpretable large language model reasoning](#). In *The Twelfth International Conference on Learning Representations*.
- Lin hao Luo, Zicheng Zhao, Gholamreza Haffari, Yuan-Fang Li, Chen Gong, and Shirui Pan. 2025. Graph-constrained reasoning: Faithful reasoning on knowledge graphs with large language models. In *International Conference on Machine Learning*, pages 41540–41565. PMLR.
- Shangwen Lv, Daya Guo, Jingjing Xu, Duyu Tang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, Guihong Cao, and Songlin Hu. 2020. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8449–8456.
- Costas Mavromatis and George Karypis. 2025. Gnnrag: Graph neural retrieval for efficient large language model reasoning on knowledge graphs. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 16682–16699.
- Frans A Oliehoek, Christopher Amato, and 1 others. 2016. *A concise introduction to decentralized POMDPs*, volume 1. Springer.
- Chanwoo Park, Seungju Han, Xingzhi Guo, Asuman E Ozdaglar, Kaiqing Zhang, and Joo-Kyung Kim. 2025. Maporl: Multi-agent post-co-training for collaborative large language models with reinforcement learning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 30215–30248.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. 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 3982–3992.
- Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.
- Xiangqing Shen, Fanfan Wang, Zinong Yang, Bing Wang, Wenli Du, Chengqing Zong, and Rui Xia. 2026. Reason-align-respond: Aligning llm reasoning with knowledge graphs for kgqa. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Haitian Sun, Tania Bedrax-Weiss, and William Cohen. 2019. Pullnet: Open domain question answering with iterative retrieval on knowledge bases and text. 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 2380–2390.
- Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Lionel Ni, Heung-Yeung Shum, and Jian Guo. 2024. [Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph](#). In *The Twelfth International Conference on Learning Representations*.
- Xingyu Tan, Xiaoyang Wang, Qing Liu, Xiwei Xu, Xin Yuan, and Wenjie Zhang. 2025. Paths-over-graph: Knowledge graph empowered large language model reasoning. In *Proceedings of the ACM on Web Conference 2025*, pages 3505–3522.
- Qwen Team. 2025. [Qwen2.5 technical report](#). Preprint, arXiv:2412.15115.
- Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O’Sullivan, and Hoang D Nguyen. 2025. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*.
- Song Wang, Junhong Lin, Xiaojie Guo, Julian Shun, Jundong Li, and Yada Zhu. 2025a. [Reasoning of large language models over knowledge graphs with super-relations](#). In *The Thirteenth International Conference on Learning Representations*.
- Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in neural information processing systems*, 33:5776–5788.
- Zihan Wang, Kangrui Wang, Qineng Wang, Pingyue Zhang, Linjie Li, Zhengyuan Yang, Xing Jin, Kefan Yu, Minh Nhat Nguyen, Licheng Liu, and 1 others. 2025b. Ragen: Understanding self-evolution in llm agents via multi-turn reinforcement learning. *arXiv preprint arXiv:2504.20073*.
- Yuan Wei, Xiaohan Shan, Ran Miao, and Jianmin Li. 2025. Lero: Llm-driven evolutionary framework with hybrid rewards and enhanced observation for multi-agent reinforcement learning. In *International Conference on Intelligent Computing*, pages 15–26. Springer.
- Jiasheng Xu, Mingda Li, Yongqiang Tang, Peijie Wang, and Wensheng Zhang. 2026. Towards open-world retrieval-augmented generation on knowledge graph: A multi-agent collaboration framework. In *Proceedings of the ACM Web Conference 2026*, pages 2171–2182.
- Yao Xu, Shizhu He, Jiabei Chen, Zihao Wang, Yangqiu Song, Hanghang Tong, Guang Liu, Jun Zhao, and Kang Liu. 2024. Generate-on-graph: Treat llm as

- both agent and kg for incomplete knowledge graph question answering. In *Proceedings of the 2024 conference on empirical methods in natural language processing*, pages 18410–18430.
- Bingcong Xue and Lei Zou. 2022. Knowledge graph quality management: a comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4969–4988.
- Wei Yang and Jesse Thomason. 2026. Learning to deliberate: Meta-policy collaboration for agentic llms with multi-agent reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 29820–29828.
- Chao Yu, Akash Velu, Eugene Vinitzky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624.
- Guibin Zhang, Kaijie Chen, Guancheng Wan, Heng Chang, Hong Cheng, Kun Wang, Shuyue Hu, and Lei Bai. 2025a. Evoflow: Evolving diverse agentic workflows on the fly. *arXiv preprint arXiv:2502.07373*.
- Jiayi Zhang, Jinyu Xiang, Zhaoyang Yu, Fengwei Teng, Xiong-Hui Chen, Jiaqi Chen, Mingchen Zhuge, Xin Cheng, Sirui Hong, Jinlin Wang, Bingnan Zheng, Bang Liu, Yuyu Luo, and Chenglin Wu. 2025b. [AFlow: Automating agentic workflow generation](#). In *The Thirteenth International Conference on Learning Representations*.
- Dongzhuoran Zhou, Yuqicheng Zhu, Xiaxia Wang, Hongkuan Zhou, Yuan He, Jiaoyan Chen, Steffen Staab, and Evgeny Kharlamov. 2026. What breaks knowledge graph based rag? benchmarking and empirical insights into reasoning under incomplete knowledge. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2522–2538.

A Algorithm

Algorithm 1: Intent-Driven Logical Compensation (IDLC)

Require: Question q , Knowledge Graph G , External Knowledge $\mathcal{D}$

Ensure: Predicted Answer Set $\mathcal{Y}_{ans}$

```

1: Initialize replay buffer  $\mathcal{B} \leftarrow \emptyset$  and predicted
   answer set  $\mathcal{Y}_{ans} \leftarrow \emptyset$ ;
2: Form anchor set  $\mathcal{E}_{anc}$  via Eq. (1);
3: for each training episode do
4:   Initialize parallel trajectories starting from
     each  $e \in \mathcal{E}_{anc}$  ( $e_1 = e$ );
5:   for  $t = 1$  to  $T$  do
6:      $o_{i,t} = \mathcal{O}(s_t, i)$  for agent  $i \in \{x, r, v\}$ ;
7:      $\mathbf{a}_t = (a_{x,t}, a_{r,t}, a_{v,t}) \sim \pi_{\theta_i}$ ;
8:     Evaluate  $g_t$  via Eq. (2);
9:     if  $g_t = 1$  then
10:      Retrieve  $\mathbf{k}_{ext}$  from  $\mathcal{D}$  via Eq. (3);
11:     end if
12:     Calculate  $F_t$  via Eq. (4);
13:     if  $F_t < \theta(s_t)$  then
14:       Routing Policy: Prune path  $p_t$  and
        Reroute;
15:     else
16:       Update  $\mathcal{Y}_{ans} \leftarrow \mathcal{Y}_{ans} \cup \{e_t\}$  via
        Eq. (5);
17:       break; {Early stopping}
18:     end if
19:     if  $t = T$  and  $e_t \notin \mathcal{Y}_{ans}$  then
20:       Update  $\mathcal{Y}_{ans} \leftarrow \mathcal{Y}_{ans} \cup \{e_t\}$  via
        Eq. (5);
21:     end if
22:     Collect  $R_t$  via Eq. (6) and store transition
        into  $\mathcal{B}$ ;
23:   end for
24:   Update  $\pi_{\theta_i}$  and  $V_\phi$  via Eq. (7)–(8), then clear
      $\mathcal{B} \leftarrow \emptyset$ ;
25: end for
26: return  $\mathcal{Y}_{ans}$ 

```

B Dataset and Evaluation

B.1 Dataset Construction and Statistics

IKG-CWQ/WebQSP. These datasets consist of 1,000 questions sampled from the original CWQ and WebQSP benchmarks. For each dataset, four degrees of incomplete KGs are generated by randomly removing 20%, 40%, 60%, or 80% crucial triples, which appear in the gold relation path. These variants, denoted as IKG-

20%/40%/60%/80%, are used to assess model robustness under varying degrees of KG incompleteness. The average number of deleted edges for each question is detailed in Table 5, while Table 6 provides statistical insights into the topic entities.

Dataset	IKG-20%	IKG-40%	IKG-60%	IKG-80%
CWQ	2.2	4.3	6.4	7.9
WebQSP	6.6	13.9	20.3	27.4

Table 5: Average number of edges deleted per question under different incompleteness degrees.

Dataset		IKG-20%	IKG-40%	IKG-60%	IKG-80%
CWQ	Median number of neighbor nodes	27	26	27	27
	Number of isolated topic entities	19	42	59	53
WebQSP	Median number of neighbor nodes	428	427	427	426
	Number of isolated topic entities	1	2	1	2

Table 6: Statistics of topic nodes in Incomplete CWQ and WebQSP.

BRINK Benchmark. The BRINK benchmark (Zhou et al., 2026) systematically evaluates reasoning capacity by mining high-confidence Horn rules (e.g., $B \Rightarrow H$) via the AMIE3 algorithm. For each grounded rule, the head triple H is removed while the supporting body atoms B are preserved. This ensures that the target answer cannot be retrieved through direct lookup but remains logically inferable through alternative reasoning paths. We utilize the official **Train/Val/Test** splits to optimize and evaluate IDLC. Comprehensive statistics for the BRINK datasets are summarized in Table 7.

Dataset	#Triples	Train	Val	Test	Total Qs
Family-Complete	17,615	1,749	218	198	2,165
Family-Incomplete	15,785	1,749	218	198	2,165
FB15k-237-Complete	204,087	4,374	535	540	5,449
FB15k-237-Incomplete	198,183	4,374	535	540	5,449
Wikidata5m-Complete	20,510,107	27,720	3,466	3,465	34,651
Wikidata5m-Incomplete	20,478,006	27,720	3,466	3,465	34,651

Table 7: Statistics of the BRINK datasets across complete and incomplete settings.

B.2 Evaluation Metrics

To provide a standardized and rigorous assessment of model performance, we employ the following metrics:

- **Hits@1 (Exact Match):** This metric calculates the proportion of queries for which the predicted answer set $\mathcal{P}_q$ matches the ground-

truth set $\mathcal{A}_q$ exactly:

$$\text{Hits@1} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \mathbb{I}(\mathcal{P}_q = \mathcal{A}_q), \quad (9)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

- **F1-score:** The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation of both coverage and correctness:

$$\text{F1} = \frac{1}{|\mathcal{Q}|} \sum_{q \in \mathcal{Q}} \frac{2 \cdot |\mathcal{P}_q \cap \mathcal{A}_q|}{|\mathcal{P}_q| + |\mathcal{A}_q|}. \quad (10)$$

C Baselines Details

In this section, we provide detailed descriptions of the representative baselines categorized by their technical paradigms.

Fine-tuned Methods

- **ChatKBQA** (Luo et al., 2024): A semantic parsing-based framework that fine-tunes LLMs to map natural language questions into executable SPARQL queries.
- **RoG** (LUO et al., 2024): A retrieval-guided method that fine-tunes LLMs to perform step-by-step reasoning along pre-retrieved KG paths, ensuring generated answers are structurally grounded.
- **GNN-RAG** (Mavromatis and Karypis, 2025): A hybrid framework that combines GNN-based structural reasoning with LLM-based language understanding by extracting KG reasoning paths from dense subgraphs.

Symbolic Exploration Frameworks

- **ToG** (Sun et al., 2024): A training-free framework that employs LLMs to perform heuristic beam search over KGs, enabling explicit reasoning path tracing and self-correction.
- **PoG** (Tan et al., 2025): A method designed for multi-hop and multi-entity queries that integrates dynamic reasoning paths and prunes irrelevant noise using graph structural filters.
- **ReKnoS** (Wang et al., 2025a): A framework that introduces super-relations to summarize and connect various relational paths, enabling forward and backward reasoning to reduce the non-retrieval rate.

Collaborative Reasoning Agents

- **GoG** (Xu et al., 2024): A framework for IKGQA that utilizes a “Thinking-Searching-Generating” workflow to infer missing triples by combining KG subgraphs with LLM parametric knowledge.
- **AnchorRAG** (Xu et al., 2026): A multi-agent collaboration framework that grounds reasoning on reliable “anchor” entities and utilizes specialized agents to coordinate graph exploration and external retrieval.
- **DoM** (Liu et al., 2026a): A multi-agent debate framework that iteratively reconciles structured KG paths with unstructured evidence through multi-turn dialogue to mitigate knowledge incompleteness.

D Implementation Details

Inference and Generation. We utilize the HuggingFace Transformers library with 8-bit quantization and KV-caching enabled to optimize inference throughput across multiple backbones. A unified decoding configuration is applied: nucleus sampling with $p = 0.95$, a maximum output length of 512 tokens, and a candidate evidence pool size $|\mathcal{K}_t| = 3$ for the RAG agent. To balance exploration and stability, the temperature is set to 0.4 during the grounding and exploration stages to facilitate diverse evidence gathering, and reduced to 0.0 for final reasoning to ensure deterministic outputs.

Policy Optimization (MAPPO). IDLC is optimized using the MAPPO algorithm on four NVIDIA A100 GPUs (80GB) with mixed precision (bfloat16). We apply LoRA to the Query, Value, and Output projections of the self-attention layers ($r = 16$, $\alpha = 32$, dropout = 0.05). Training is performed using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$) with an initial learning rate of 5×10^{-5} and a cosine decay schedule. To stabilize the multi-agent training process, gradient norms are clipped at 1.0. We incorporate an entropy bonus of 0.01 to encourage policy exploration and a KL regularization weight of 0.1 to maintain alignment with the base LLM’s reasoning capabilities. Key MARL hyperparameters include a clipping parameter $\epsilon = 0.2$, a value loss weight $\lambda = 0.5$, and a discount factor $\gamma = 0.95$.

Graph and Retrieval Setup. The knowledge graphs (Freebase and Wikidata) are hosted on a local Virtuoso (Erling and Mikhailov, 2009) server. In the semantic scoring step (Eq. 1), the top- $k = 5$ relations in the entity neighborhood are used to compute the anchor scores. For the RAG component, the retrieved Wikipedia articles are segmented into 500-token chunks. The bias of the gap indicator $\eta(s_t)$ network is initialized to 0.5, providing a neutral starting prior for the agents to learn the optimal timing for logical compensation. All results are reported as the average over fixed random seeds: 25, 42, and 99.

E Case Studies on Intent Validity and Corrective Reasoning

To qualitatively demonstrate the effectiveness of the *Explorer* in bridging structural gaps and the *Verifier* in guiding path correction, we provide two representative cases. Table 8 compares the reasoning traces of IDLC with the competitive baseline, DoM. These cases highlight how IDLC leverages Gap Descriptors to navigate isolated entities and utilizes fidelity-based feedback ($F_t < \theta(s_t)$) to prune fallacious branches.

F Impact of Varying Graph Incompleteness

To evaluate IDLC’s generalization against structural degradation, we test its performance on IKG-CWQ/WebQSP by varying triple deletion rates from 0% to 80%. Crucially, we adopt a zero-shot transfer setting, applying policies trained exclusively on Wikidata5m-Incomplete to these new domains without any additional fine-tuning. As shown in Figure 4, IDLC demonstrates superior resilience with a more graceful degradation curve than all competitive baselines. Specifically, the performance gap between IDLC and the second-best method (DoM) on IKG-CWQ widens from 2.1% at 0% deletion to 4.2% at the extreme 80% incompleteness level. This expanding margin confirms that our intent-driven deliberation effectively mitigates the structural cold-start problem. Unlike DoM which remains inherently bounded by local topological connectivity, IDLC leverages explicit Gap Descriptors to trigger precise logical compensation even when internal graph structures are nearly absent. This adaptive capability validates that IDLC successfully learns universal reasoning patterns capable of generalizing across varied do-

main and incompleteness levels.

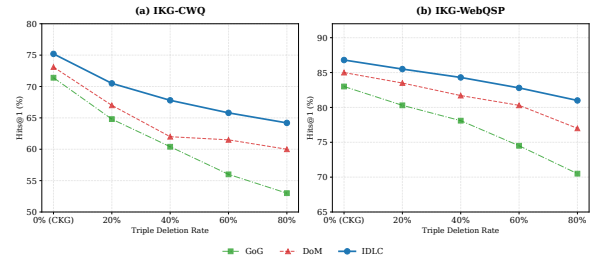


Figure 4: Performance (Hits@1) comparison under varying graph incompleteness levels in a zero-shot transfer setting.

G Hyperparameter Sensitivity

To evaluate the hyperparameter robustness of IDLC, we conduct sensitivity analyses using Qwen2.5-7B-Instruct on three key parameters governing the agents’ behaviors: candidate evidence pool size $|\mathcal{K}_t|$ (resource acquisition), maximum reasoning depth T (search boundary), and discount factor γ (strategic planning). As illustrated in Figure 5, IDLC exhibits consistent performance trends across both datasets, aligning with our design intuition. Specifically, for the candidate evidence pool size, Hits@1 increases significantly as $|\mathcal{K}_t|$ grows to 3 and plateaus thereafter, confirming that a compact set of intent-driven evidence is sufficient to bridge structural gaps while avoiding redundant semantic noise. Regarding reasoning depth, $T = 4$ emerges as the optimal boundary. Smaller values fail to capture complex multi-hop dependencies, whereas $T \geq 5$ leads to slight degradation due to cumulative logic-misalignment and hallucinated paths in incomplete structures. Finally, the monotonic gain as γ increases to 0.95 validates our DecPOMDP formulation, as a higher discount factor incentivizes the centralized critic V_ϕ to prioritize long-term logical fidelity over short-sighted symbolic exploration.

H Error Analysis

To investigate the performance boundaries of IDLC, we manually categorize 200 failed trajectories from each dataset, as visualized in Figure 6. This cross-dataset taxonomy reveals that failure modes are deeply rooted in the intrinsic topological properties of the KGs. In the large-scale Wikidata5m-Incomplete, the primary bottleneck is *Search Deadlock* (42%), stemming from the extreme structural fragmentation that hinders

Component	Case 1: Sequential Logical Bridge	Case 2: Multi-Constraint Disambiguation
Question	Where did the famous person who went to Al-Thager Model School get killed?	What NFL Team did Peyton Manning’s father play for also won the 2010 NFC Championship Game?
KG Status	Structural Void: Missing <i>education.institution</i> and <i>place_of_death</i> relations; school node is isolated.	Broken Constraint: Missing <i>parents</i> relation. Connection between Peyton and Archie Manning is severed.
DoM Search (Baseline)	Trapped at <i>Al-Thager Model School</i> ; fails to bridge to the target person due to missing edges.	Retrieves teams associated with <i>Peyton Manning</i> , ignoring parental constraints due to link breakage.
IDLC Intent (Explorer)	“Identify the notable alumnus of Al-Thager Model School and retrieve their location of death.”	“Identify the father of Peyton Manning and the NFL team he played for that won the 2010 NFC.”
Reasoning Trace	1. KG-only: Fails to find paths ($F_t < \theta(s_t)$). 2. Corrective: Triggers RAG based on intent. 3. Verification: Found “Osama bin Laden” ($F_t = 0.94$).	1. LLM-Prior: Confused by Peyton’s own teams ($F_t = 0.45$). 2. Corrective: Prunes path due to $F_t < \theta(s_t)$, focuses on father’s intent. 3. Verification: Found “Archie Manning” ($F_t = 0.91$).
Evidence Support	<i>Verified Fact:</i> “Osama bin Laden, an alumnus of Al-Thager, was killed in Abbottabad.”	<i>Verified Fact:</i> “Archie Manning’s son Peyton... Archie played for the Saints, who won the 2010 NFC.”
Final Answer	Abbottabad	New Orleans Saints

Table 8: Comparison of reasoning traces between DoM and IDLC. The *Reasoning Trace* highlights IDLC’s ability to prune low-fidelity paths ($F_t < \theta(s_t)$) and re-evaluate symbolic branches via the deliberative loop.

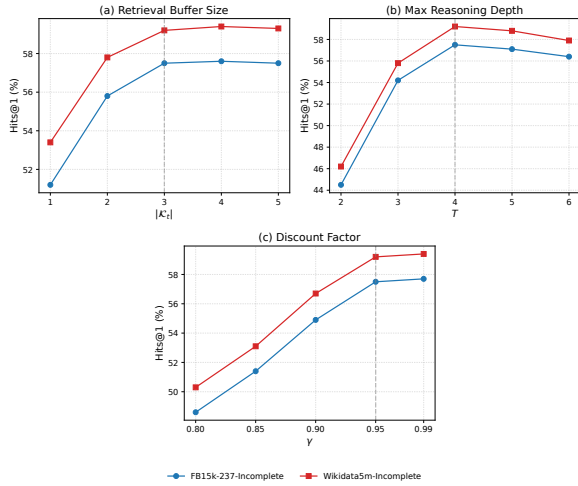


Figure 5: Hyperparameter sensitivity analysis of IDLC on FB15k-237/Wikidata5m-Incomplete.

the Explorer from formulating valid Gap Descriptors. Meanwhile, FB15k-237-Incomplete exhibits a higher propensity for *Verification Over-pruning* (32%) and *Knowledge Absence* (30%). Despite its connectivity, the dense and semantically overlapping relation set in FB15k-237 complicates logical

verification, while its specialized long-tail facts are less prevalent in external corpora. For the logically rigid Family-Incomplete, errors are dominated by *Knowledge Absence* (45%) and *Grounding Error* (33%), reflecting a closed-world nature where missing private triples are irrecoverable and entity disambiguation is challenged by naming similarities. These findings underscore that while IDLC effectively repairs fractured logic, future efforts should focus on enhancing structural resilience in incomplete environments and refining the dynamic calibration of the logic-fidelity threshold $\theta(s_t)$.

I Prompt Templates

This section presents the comprehensive prompt templates that define the reasoning behavior and interaction protocols for the specialized agents within the IDLC framework. Unlike training-free paradigms, these templates serve as the *instructional grounding* for our MARL policy optimization. They establish the initial semantic action space for the Dec-POMDP, ensuring that the generated intents and fidelity assessments remain logi-

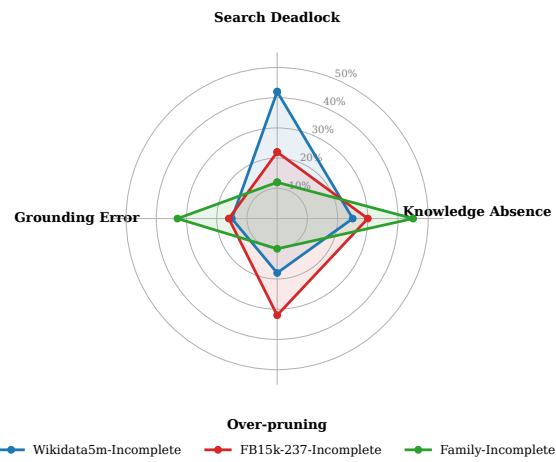


Figure 6: Manual error taxonomy and distribution across three IKGQA datasets.

cally consistent and interpretable during the joint optimization process.

Active KG Explorer: Logical Gap Identification Agent

Instruction: You are an expert Knowledge Graph explorer acting in a incomplete graph environment. Your goal is to identify why a symbolic reasoning path has reached a structural void and describe the “logical vacancy” required to reach the target answer.

Input Context:

- Goal Question: <Question>
- Current Path History: <Path_History>
- Current Anchor Entity: <Current_Entity>

Task Requirement: Generate a concise **Gap Descriptor** (D_{gap}) that defines the missing logical connection (e.g., “Identify the spouse of [Entity] who is also a professional athlete”). Do not attempt to guess the final answer; focus exclusively on characterizing the structural deficiency.

A:

Thinking: <Internal_Logic_Analysis>

Gap Descriptor (D_{gap}): <Extracted_Logical_Intent>

RAG Agent: Intent-Driven Question Reformulation

Instruction: You are a strategic information retriever. This module is triggered when the Explorer agent x evaluates the Gap Indicator $g_t = 1$. Based on the “Gap Descriptor” (D_{gap}) provided by the KG Explorer, reformulate the retrieval question to bridge the structural breakage.

Input Context:

- Logic Gap (D_{gap}): <D_gap_output_from_Explorer>
- Global Reasoning Context: <Full_Question_and_Context>

Task Requirement: Your reformulated question (q^*) must be highly targeted to find external evidence that specifically satisfies the logical intent. Use the gap description to anchor the search in external knowledge source $\mathcal{D}$ to avoid general keyword ambiguity.

A:

Thinking: <Search_Strategy_Analysis>

Search Question (q^*): <Optimized_Keyword_Question>

Logical Consistency Verifier: Step-wise Fidelity Assessment

Instruction: You are a logical consistency auditor. Your task is to evaluate whether a proposed reasoning hop is factually supported by the retrieved external evidence. This assessment provides the basis for the *Fidelity Score* (F_t).

Input Context:

- Proposed Symbolic Step: ($\langle \text{Subject} \rangle \xrightarrow{\text{Relation}} \langle \text{Object} \rangle$)
- External Evidence: <Retrieved_Evidence_Passage>

Task Requirement: Perform a rigorous cross-validation. Determine if the evidence explicitly or implicitly confirms this connection. Output a numerical **Fidelity Score** (F_t) between 0.0 (Pure Hallucination) and 1.0 (Strictly Grounded).

A:

Thinking: <Evidence_Alignment_Analysis>

Fidelity Score (F_t): <Numerical_Value_for_Reward_Calculation>

Routing Policy: Reasoning Trajectory Redirection

Instruction: The previous reasoning trajectory has been flagged for low logical fidelity ($F_t < \theta(s_t)$). You must now identify the specific point of semantic drift and propose a redirection strategy.

Input Context:

- Failed Reasoning Trace: <Reasoning_Log>
- Discrepancy Analysis: <Verification_Feedback>

Task Requirement: Identify the “Drift Point” (the exact hop where the logic failed) and provide a new “Instruction” to guide the Explorer toward a more reliable reasoning branch.

A:

Thinking: <Error_Propagation_Analysis>

Drift Point Identification: <Step_Number>

Redirection Strategy: <New_Search_Intent_to_Reinitialize_Path>