



ICLR
2024



MONASH
University



Reasoning on Graphs: Faithful and Interpretable Large Language Models Reasoning

Linhao Luo, Yuan-Fang Li, Gholamreza Haffari, Shirui Pan

Monash University and Griffith University

Presenter: Linhao Luo



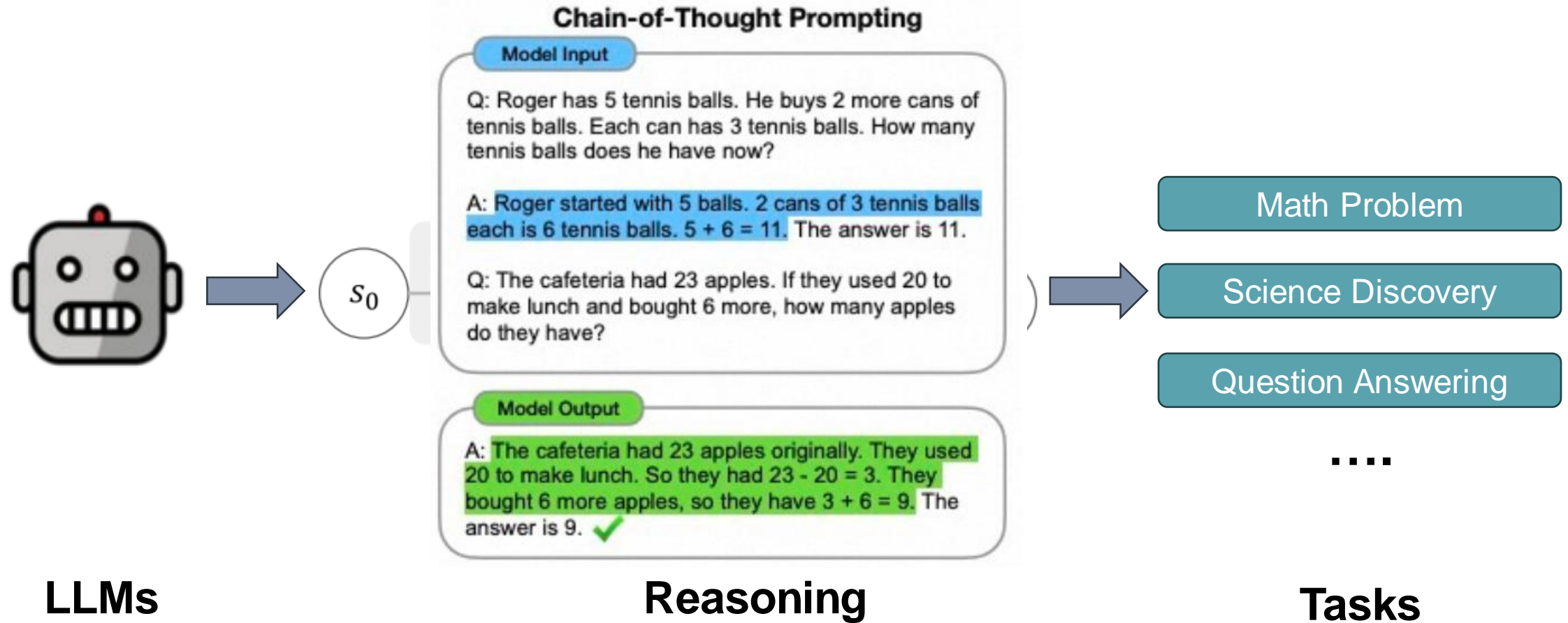
Paper



Code

Background

- Large language models (LLMs) exhibit a great reasoning ability.



Limitations

- LLMs **lack up-to-date knowledge** and experience **hallucinations** during reasoning.

Question

What product did Apple release in 2023?

Output

Sorry, **I do not have knowledge** after Sept. 2021.
Could you provide some additional information?

Lack of Knowledge

Question

Who is the brother of Justin Bieber

Output

Justin Bieber is the child of Jeremy Bieber, who **has a daughter** named Allie Bieber. Thus, the brother of Justin Bieber is **Allie Bieber**.

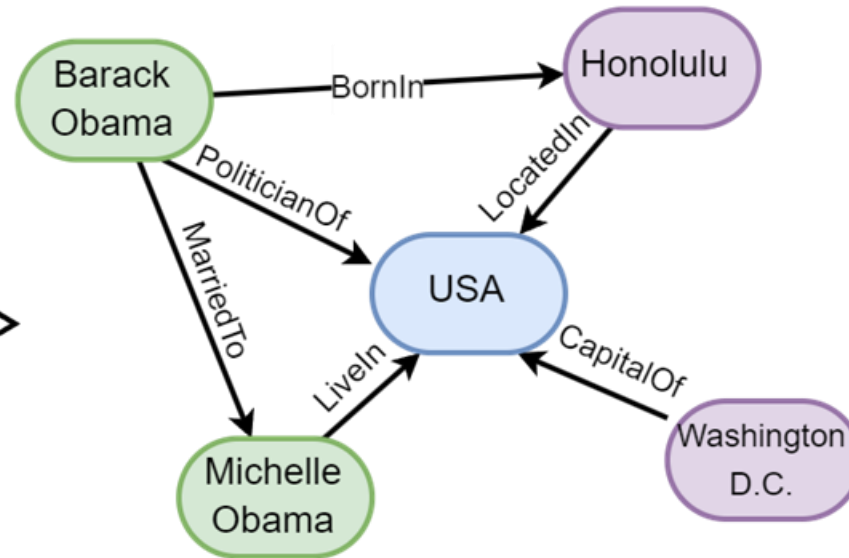
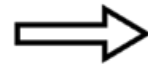
Hallucination

Motivations

- Knowledge graphs (KGs) contain abundant factual knowledge in a **structured format**.



Wikipedia



Motivations

- Knowledge graphs (KGs) offer a **reliable source of knowledge for reasoning**.

Question

What product did Apple release in 2023?



Output

Sorry, **I do not have knowledge** after Sept. 2021.
Could you provide some additional information?

Lack of Knowledge

Factual Knowledge ↑

Triple: (Iphone 15, released_at, 2023)

Question

Who is the brother of Justin Bieber



Output

Justin Bieber is the child of Jeremy Bieber, who
has a daughter named Allie Bieber. Thus, the
brother of Justin Bieber is **Allie Bieber**.

Hallucination

Reasoning Guidance ↑

Relation path: child_of → has_son

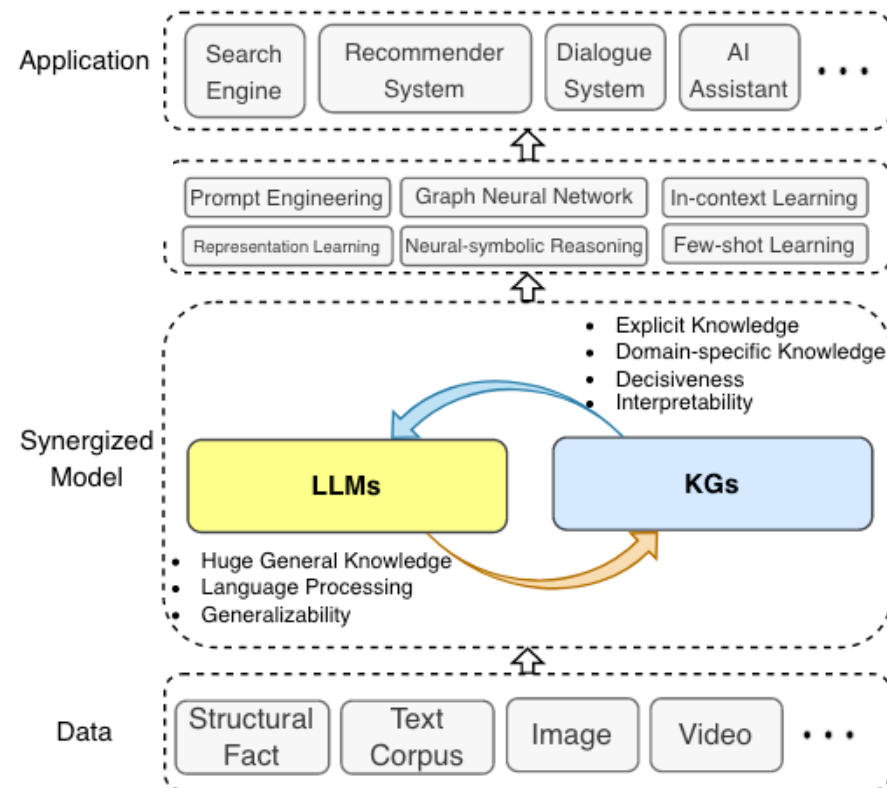
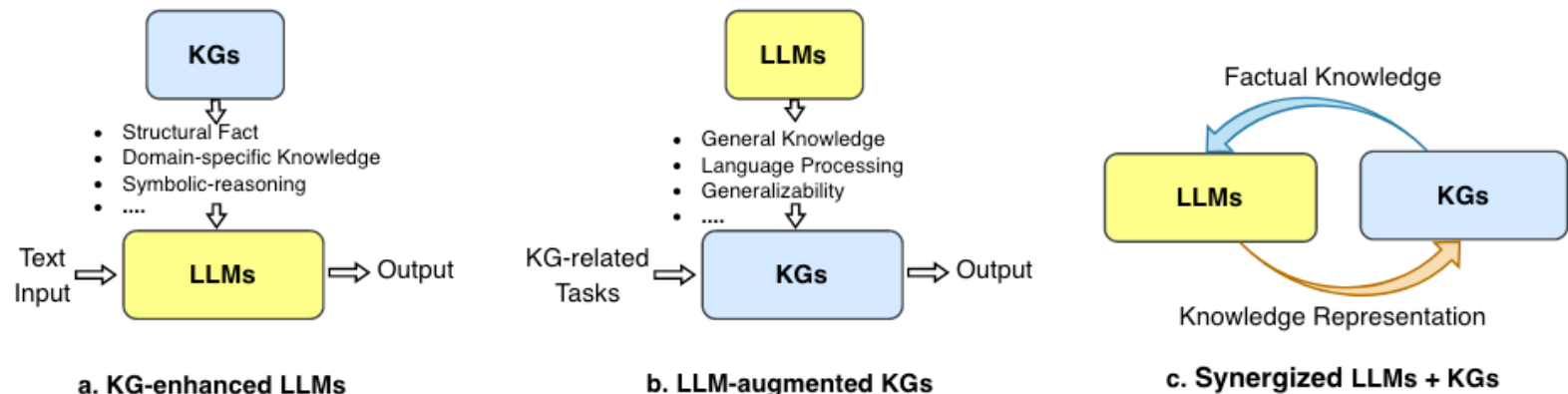
KGs + LLMs

• Unifying Large Language Models and Knowledge Graphs: A Roadmap



Shirui Pan, *Senior Member, IEEE*, Linhao Luo,
Yufei Wang, Chen Chen, Jiapu Wang, Xindong Wu, *Fellow, IEEE*

TKDE



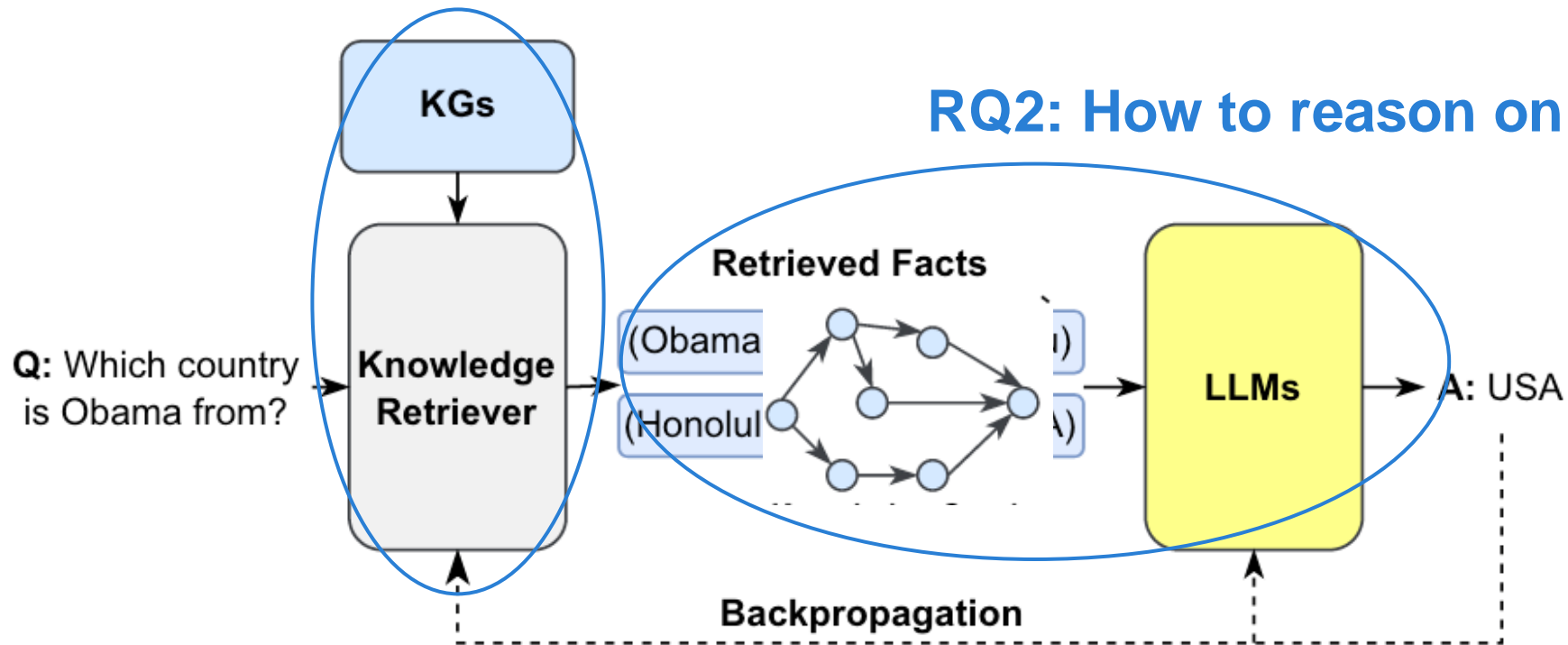
The general roadmap of unifying KGs and LLMs. (a.) KG-enhanced LLMs. (b.) LLM-augmented KGs. (c.) Synergized LLMs + KGs.

Fig. 7. The general framework of the *Synergized LLMs + KGs*, which contains four layers: 1) *Data*, 2) *Synergized Model*, 3) *Technique*, and 4) *Application*.

KG-enhanced LLM Reasoning

- Retrieval-Augmented LLM Reasoning (RAG)

RQ1: How to effectively retrieve from KGs?

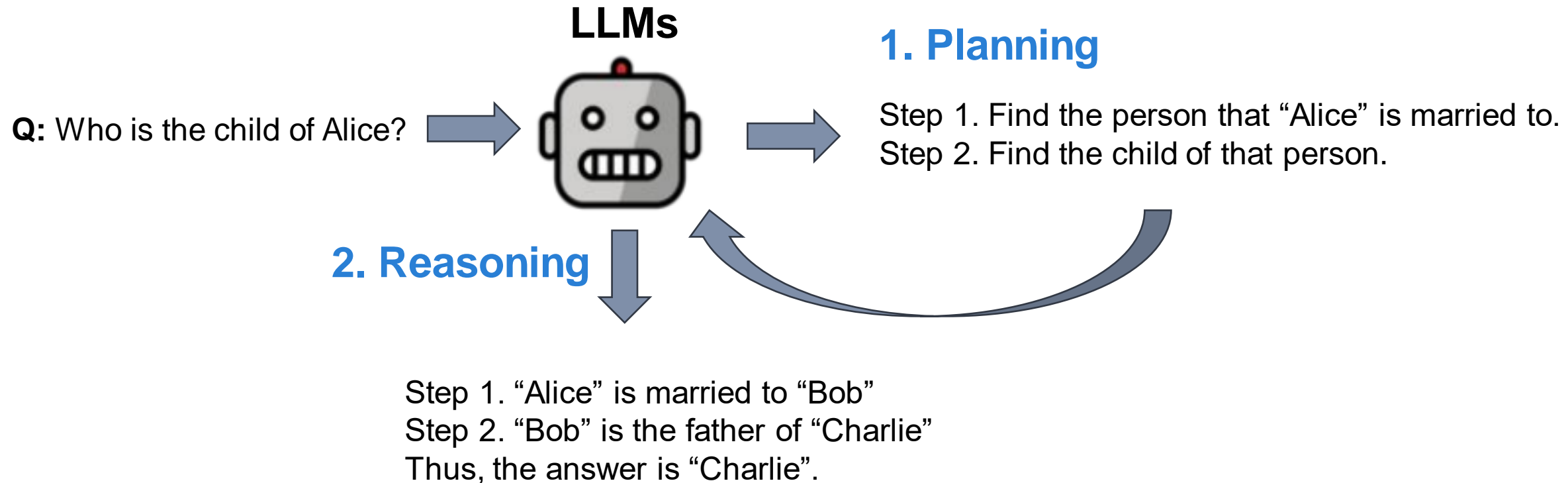


How to effectively utilize the KGs for LLMs reasoning?

Challenges: How to reason on graphs?

- **Plan-and-solve reasoning**

- The plan is a hidden logic that can guide the reasoning.



Challenges: How to reason on graphs?

- **Relation paths as plans**

- Relation paths are a **sequence of relations** that can serve as faithful plans for reasoning on graphs.

- **Example:**

- **Question:**

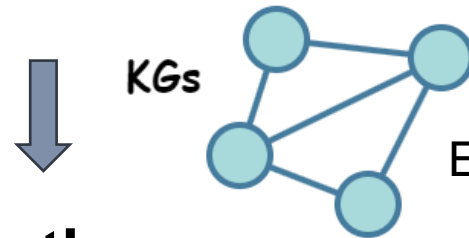
- Who is the child of Alice?

- **Relation path z :**

$z = \text{marry_to} \rightarrow \text{father_of}$

Plan:

- Step 1. Find the person that “Alice” is married to.
Step 2. Find the child of that person.



Execute the plan on KGs to **retrieve** reasoning paths.

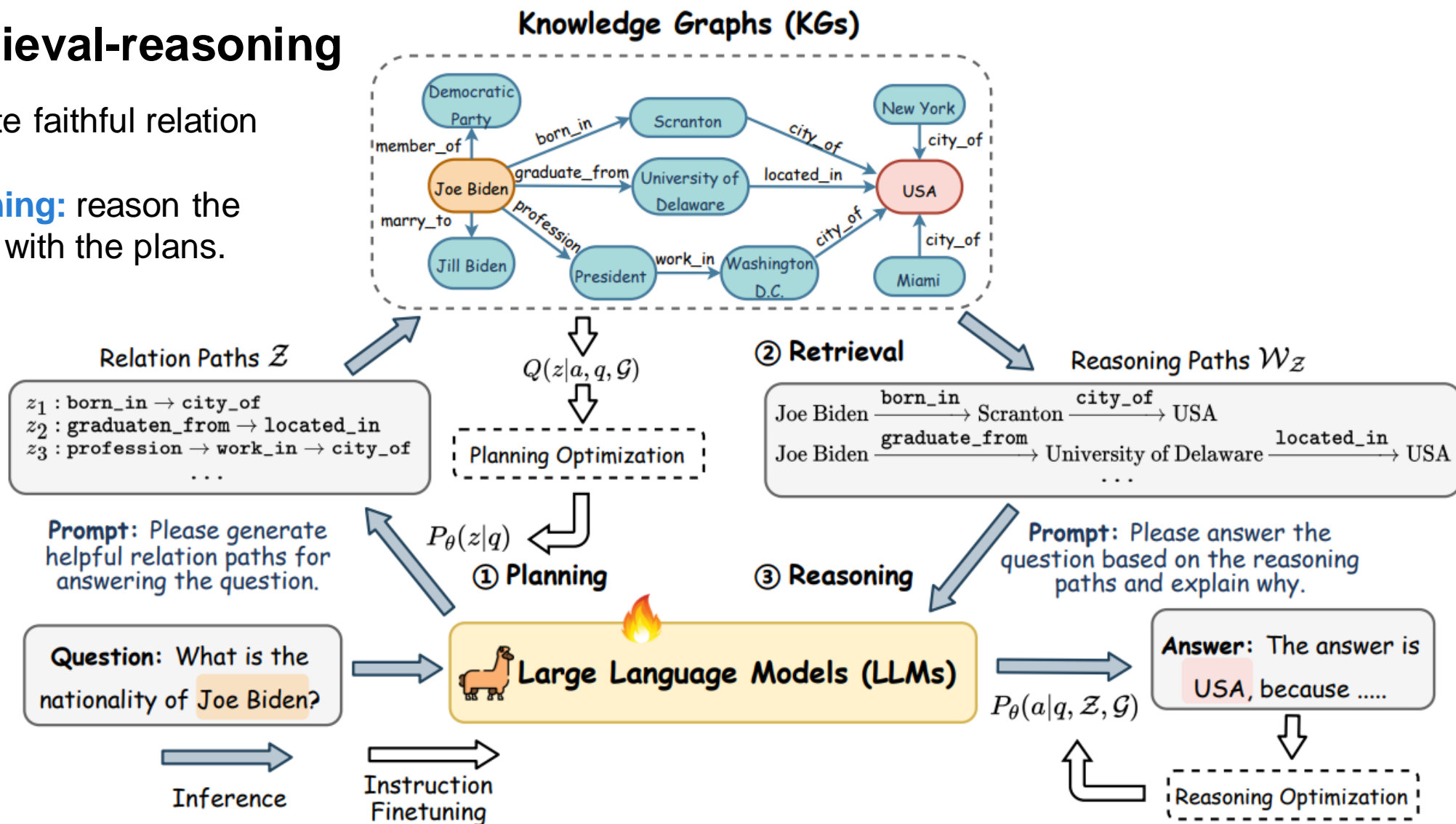
- **Reasoning paths w_z :**

$w_z = \text{Alice} \xrightarrow{\text{marry_to}} \text{Bob} \xrightarrow{\text{father_of}} \text{Charlie}$ **Answer**

Reasoning on graphs (RoG)

- **Planning-retrieval-reasoning**

1. **Planning:** generate faithful relation paths as plans.
2. **Retrieval-Reasoning:** reason the answer on graphs with the plans.



Reasoning on graphs (RoG): Planning-retrieval-reasoning

$$P_{\theta}(a|q, \mathcal{G}) = \sum_{z \in \mathcal{Z}} \overset{\text{Reasoning}}{P_{\theta}(a|q, z, \mathcal{G})} \overset{\text{Planning}}{P_{\theta}(\boxed{z}|q)},$$

↓
How to make sure the plan is faithful?

This function cannot be directly optimized as we do not know the golden plan.

Reasoning on graphs (RoG)

$$P_{\theta}(a|q, \mathcal{G}) = \sum_{z \in \mathcal{Z}} \overset{\text{Reasoning}}{P_{\theta}(a|q, z, \mathcal{G})} \overset{\text{Planning}}{P_{\theta}(z|q)},$$



ELBO loss

$$\log P(a|q, \mathcal{G}) \geq \mathbb{E}_{z \sim Q(z)} [\log P_{\theta}(a|q, z, \mathcal{G})] - D_{\text{KL}}(Q(z) \| P_{\theta}(z|q)),$$

Estimate the posterior distribution of faithful relation paths with the **shortest path** connecting question and answer entities on KGs.

$$Q(z) \simeq Q(z|a, q, \mathcal{G}) = \begin{cases} 1, & \exists w_z(e_q, e_a) \in \mathcal{G}, \\ 0, & \text{else.} \end{cases}$$



Two instruction tuning tasks

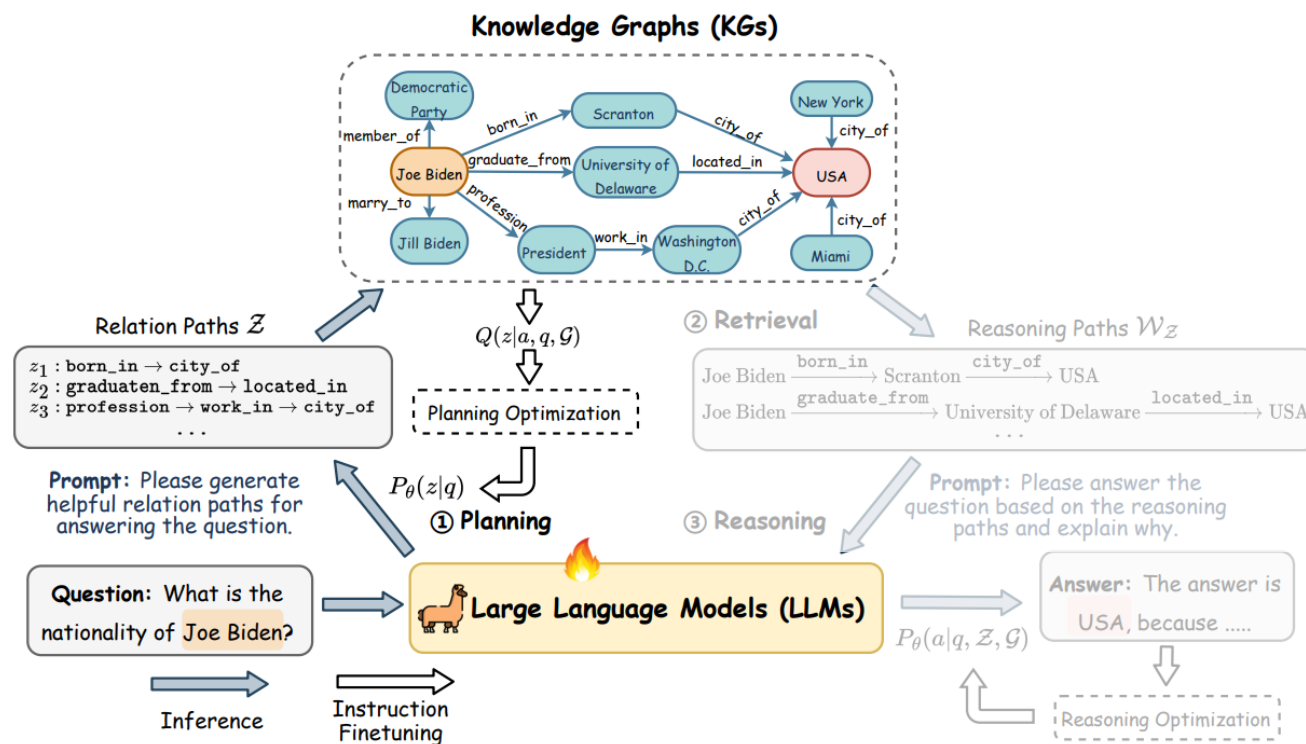
$$\mathcal{L} = \underbrace{\log P_{\theta}(a|q, \mathcal{Z}_K^*, \mathcal{G})}_{\text{Retrieval-reasoning}} + \underbrace{\frac{1}{|\mathcal{Z}^*|} \sum_{z \in \mathcal{Z}^*} \log P_{\theta}(z|q)}_{\text{Planning}}.$$

Planning-retrieval-reasoning

- **Planning:** generate faithful relation paths as plans.

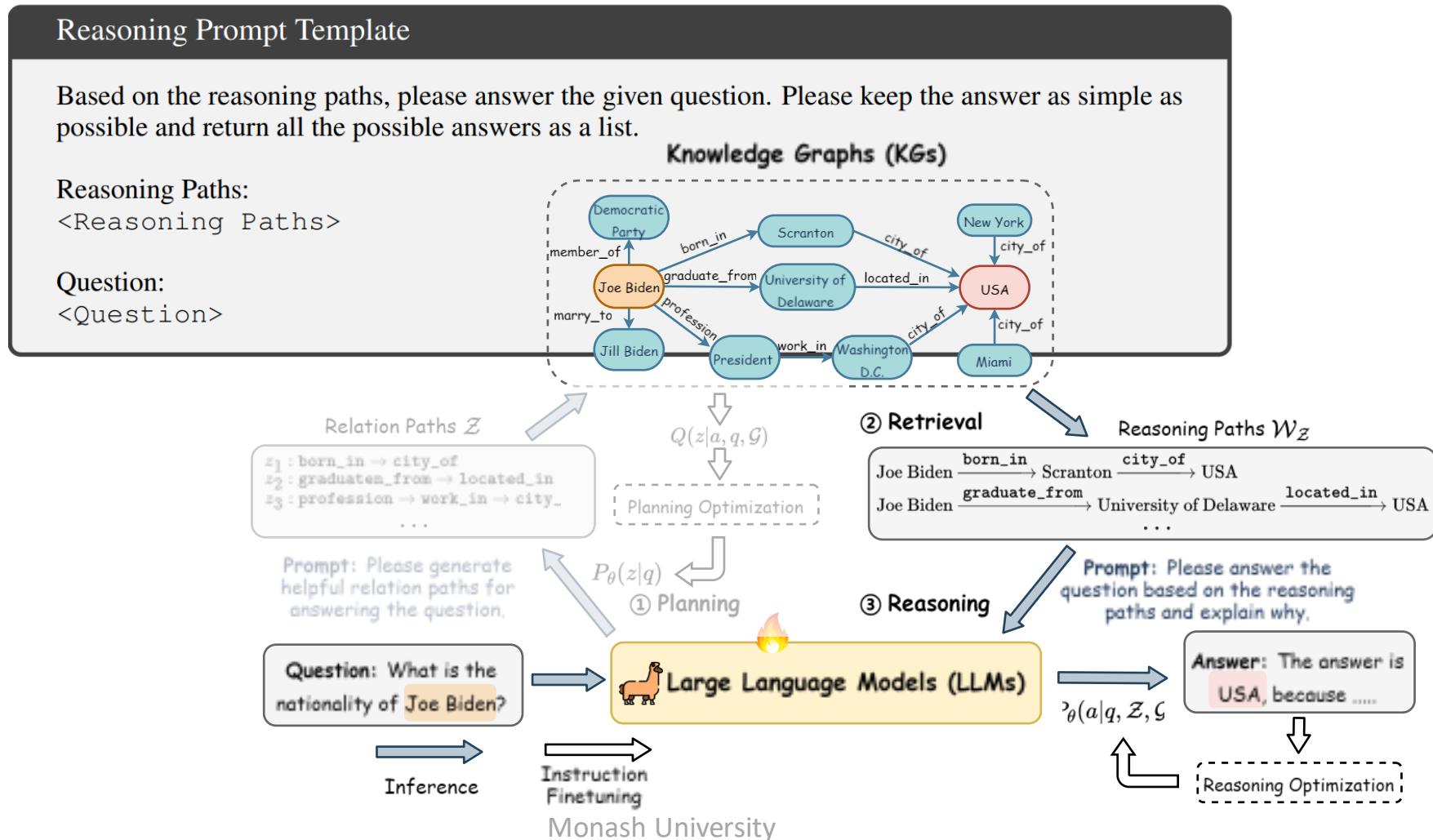
Planning Prompt Template

Please generate a valid relation path that can be helpful for answering the following question:
<Question>



Planning-retrieval-reasoning

- **Retrieval-Reasoning:** reason the answer on graphs with the plans.



Experiments

- **RQ1:** Can RoG perform effective reasoning?
- **RQ2:** Can RoG enable effective retrieval?
- **RQ3:** Can RoG be transferred to other KGs?
- **RQ4:** Can RoG generate faithful and interpretable results?

Experiments

Table 2: Performance comparison with different baselines on the two KGQA datasets.

Type	Methods	WebQSP		CWQ	
		Hits@1	F1	Hits@1	F1
Embedding	KV-Mem (Miller et al., 2016)	46.7	34.5	18.4	15.7
	EmbedKGQA (Saxena et al., 2020)	66.6	-	45.9	-
	NSM (He et al., 2021)	68.7	62.8	47.6	42.4
	TransferNet (Shi et al., 2021)	71.4	-	48.6	-
	KGT5 (Saxena et al., 2022)	56.1	-	36.5	-
Retrieval	GraftNet (Sun et al., 2018)	66.4	60.4	36.8	32.7
	PullNet (Sun et al., 2019)	68.1	-	45.9	-
	SR+NSM (Zhang et al., 2022)	68.9	64.1	50.2	47.1
	SR+NSM+E2E (Zhang et al., 2022)	69.5	64.1	49.3	46.3
Semantic Parsing	SPARQL (Sun et al., 2020)	-	-	31.6	-
	QGG (Lan & Jiang, 2020)	73.0	73.8	36.9	37.4
	ArcaneQA (Gu & Su, 2022)	-	75.3	-	-
	RnG-KBQA (Ye et al., 2022)	-	76.2	-	-
LLMs	Flan-T5-xl (Chung et al., 2022)	31.0	-	14.7	-
	Alpaca-7B (Taori et al., 2023)	51.8	-	27.4	-
	LLaMA2-Chat-7B (Touvron et al., 2023)	64.4	-	34.6	-
	ChatGPT	66.8	-	39.9	-
	ChatGPT+CoT	75.6	-	48.9	-
LLMs+KGs	KD-CoT (Wang et al., 2023b)	68.6	52.5	55.7	-
	UniKGQA (Jiang et al., 2022)	77.2	72.2	51.2	49.1
	DECAF (DPR+FiD-3B) (Yu et al., 2022a)	82.1	78.8	-	-
	RoG	85.7	70.8	62.6	56.2

Performance on KGQA tasks.

(RQ1: Effective reasoning)

Monash University

Table 6: Statistics of datasets.

Datasets	#Train	#Test	Max #hop
WebQSP	2,826	1,628	2
CWQ	27,639	3,531	4

We feed the RoG retrieved reasoning paths into other LLMs without training them.

Table 4: Effects of integrating the planning module of RoG with different LLMs for reasoning.

Methods	WebQSP		CWQ	
	Hits@1	Recall	Hits@1	Recall
ChatGPT	66.77	49.27	39.90	35.07
ChatGPT + RoG Planning	81.51	71.60	52.68	48.51
Alpaca-7B	51.78	33.65	27.44	23.62
Alpaca-7B + RoG Planning	56.16	74.20	44.04	38.46
LLaMA2-Chat-7B	64.37	44.61	34.60	29.91
LLaMA2-Chat-7B + RoG Planning	74.20	56.16	56.41	51.99
Flan-T5-xl	30.95	17.08	14.69	12.25
Flan-T5-xl + RoG Planning	67.87	44.93	37.81	32.57

Plug-and-play with other LLMs 17

(RQ2: Effective retrieval)

Experiments

- Transfer to Wiki-Movies KG.

Table 10: Statistics of constructed knowledge graphs.

KG	#Entities	#Relations	#Triples
Freebase	2,566,291	7,058	8,309,195
Wiki-Movie	43,234	9	133,582

Table 9: Statistics of MetaQA-3hop datasets.

Datasets	#Train	#Test	#hop
MetaQA-3hop	1,000	1,4274	3

Table 12: Performance of RoG on MetaQA-3hop.

Strategies	MetaQA-3hop	
	Hits@1	F1
RoG (train from scratch)	84.81	41.32
RoG (transfer from Freebase)	88.98	50.68

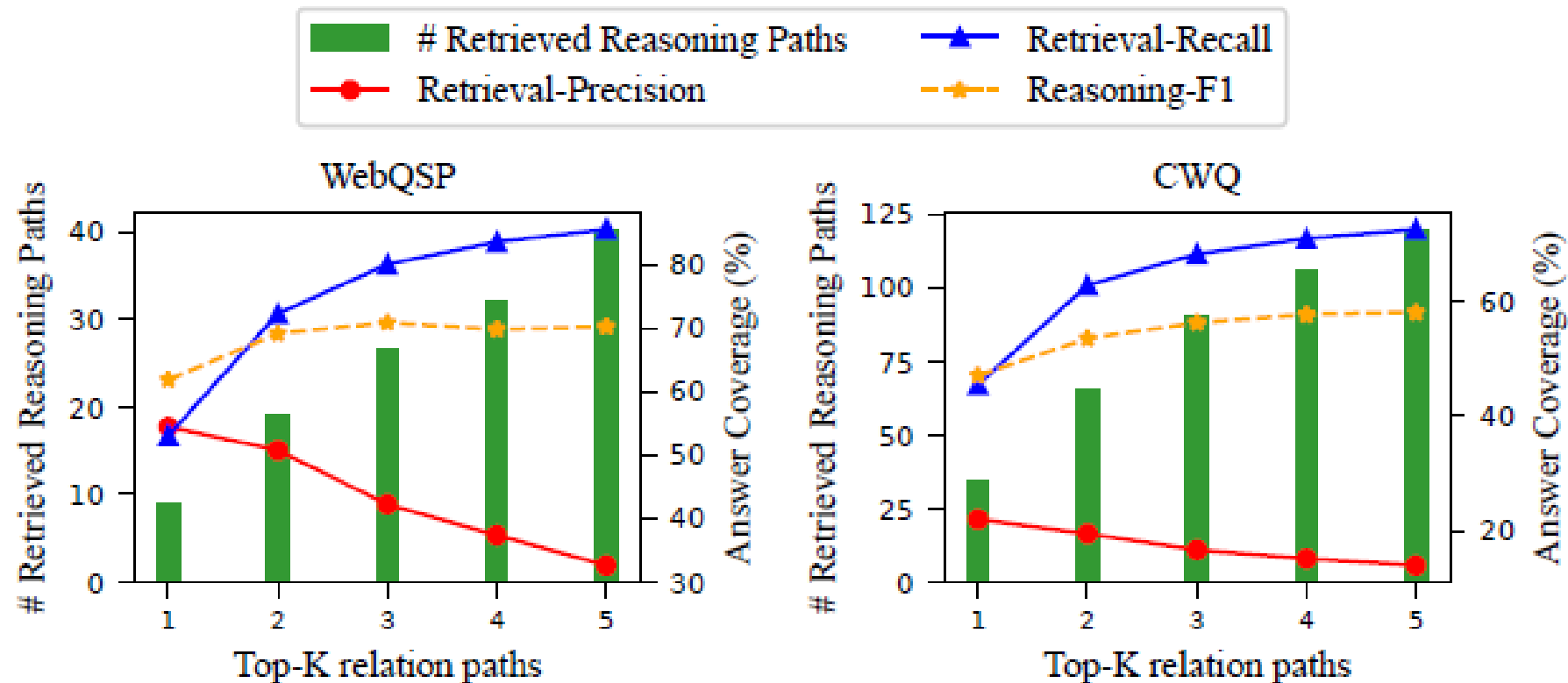
Table 13: Training time comparison.

Method	Training on Freebase	Transferring to Wiki-Movies
RoG	38 hours	2 hours

RoG can be transferred to other KGs with a few training data.
(RQ3: Transferability)

Faithfulness of generated relation Paths

- More relation paths increase answer coverage but also introduce noise, not necessarily improving reasoning performance.



A large context of LLMs cannot address all problems without a **sanitary context information**.

Faithful reasoning and interpretable results

RQ4: faithful and interpretable results

Lack of knowledge

Table 6: Example of *lack of knowledge* for ChatGPT + CoT and interpretable predictions of RoG.

Question	What type of government is used in the country with Northern District?
Answer	Parliamentary system
ChatGPT + CoT	To determine what type of government is used in a country with a Northern District, we need more specific information . The term "Northern District" does not specify the name of any particular country or provide enough context to identify its government type. Could you please provide additional details or clarify the specific country you are referring to?
Reasoning Paths	Northern District → location.administrative_division.first_level_division_of → Israel → government.form_of_government.countries → Parliamentary system
RoG	The type of government used in the country with the Northern District is a Parliamentary system .
	Explanations:
	1. Northern District → location.administrative_division.first_level_division_of → Israel: The Northern District is a part of Israel. 2. Israel → government.form_of_government.countries → Parliamentary system: Israel has a Parliamentary system of government. So, by following the reasoning paths provided, we can conclude that the country with the Northern District, which is Israel, uses a Parliamentary system of government.

Hallucination

Table 7: Example of the *hallucination* of ChatGPT + CoT and faithful reasoning of RoG.

Question	Where is the home stadium of the team who won the 1946 World Series championship?
Answer	Busch Stadium
ChatGPT + CoT	Step 1: Identify the team that won the 1946 World Series championship. The team that won the 1946 World Series championship was the St. Louis Cardinals.
	Step 2: Determine the home stadium of the St. Louis Cardinals during that time. The St. Louis Cardinals' home stadium during the 1946 season was Sportsman's Park .
	Therefore, the possible answer to the question "Where is the home stadium of the team who won the 1946 World Series championship?" is: - Sportsman's Park .
Reasoning Paths	1946 World Series → sports.sports_team.championships → St. Louis Cardinals → sports.sports_team.arena_stadium → Busch Stadium 1946 World Series → sports.sports_team.championships → St. Louis Cardinals → sports.sports_team.arena_stadium → Roger Dean Stadium
	Based on the reasoning paths provided, the home stadium of the team who won the 1946 World Series championship is Busch Stadium . This conclusion is based on the following path:
RoG	1. 1946 World Series → sports.sports_team.championships → St. Louis Cardinals → sports.sports_team.arena_stadium → Busch Stadium
	The reasoning path leads us to the St. Louis Cardinals as the team that won the 1946 World Series, and Busch Stadium is the stadium associated with the St. Louis Cardinals. Therefore, Busch Stadium is the home stadium of the team that won the 1946 World Series championship.

Thanks for your listening!



Paper



Code