NeurIPS 2025

GFM-RAG: Graph Foundation Model for Retrieval Augmented Generation

Linhao Luo[1], Zicheng Zhao[2], Gholamreza Haffari[1], Dinh Phung[1], Chen Gong[3], Shirui Pan[4]

[1]Monash University, [2]Nanjing University of Science and Technology, [3]Shanghai Jiao Tong University, [4]Griffith University

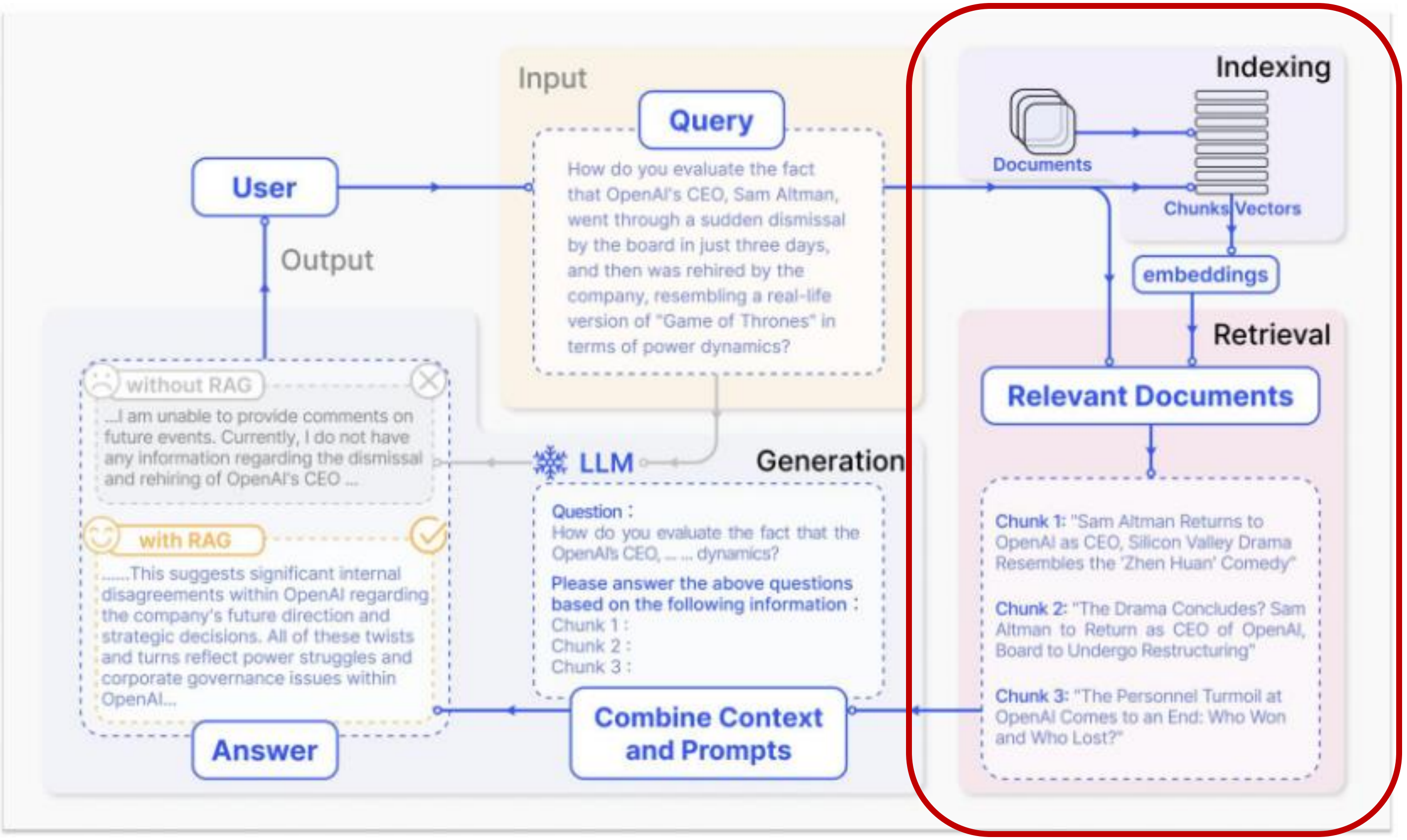
Paper

Presenter: Linhao Luo

Code

Retrieval-Augmented Generation (RAG)

- Retrieval-augmented generation (RAG) has proven effective in integrating knowledge into LLMs without training.



How do you evaluate the fact that OpenAI's CEO, Sam Altman, went through a sudden dismissal by the board in just three days, and then was rehired by the company, resembling a real-life version of "Game of Thrones" in terms of power dynamics?

...I am unable to provide comments on future events. Currently, I do not have any information regarding the dismissal and rehiring of OpenAI's CEO ...

......This suggests significant internal disagreements within OpenAI regarding the company's future direction and strategic decisions. All of these twists and turns reflect power struggles and corporate governance issues within OpenAI...

Question :
How do you evaluate the fact that the OpenAI's CEO, ... dynamics?
Please answer the above questions based on the following information :
Chunk 1 :
Chunk 2 :
Chunk 3 :

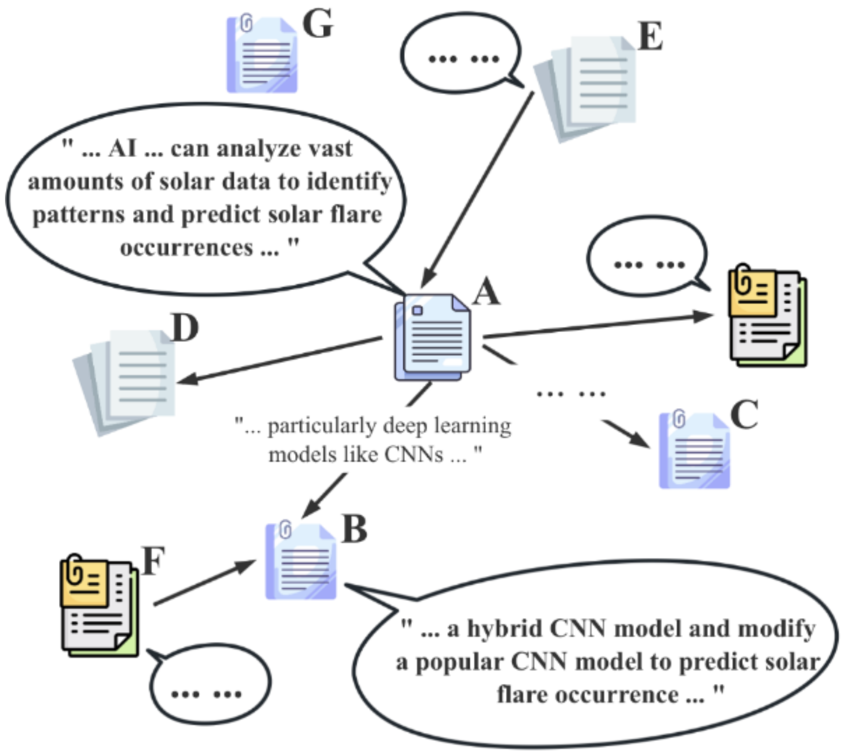
Chunk 1: "Sam Altman Returns to OpenAI as CEO, Silicon Valley Drama Resembles the 'Zhen Huan' Comedy"

Chunk 2: "The Drama Concludes? Sam Altman to Return as CEO of OpenAI, Board to Undergo Restructuring"

Chunk 3: "The Personnel Turmoil at OpenAI Comes to an End: Who Won and Who Lost?"

Limitations of RAG

- **Neglecting relationships**
  - Traditional RAGs struggle to capture complex relationships between pieces of knowledge, limiting their performance in intricate reasoning that requires integrating knowledge from multiple sources, e.g., **multi-hop reasoning.**


G
... ...
E
" ... AI ... can analyze vast amounts of solar data to identify patterns and predict solar flare occurrences ... "
A
... ...
D
... ...
"... particularly deep learning models like CNNs ... "
C
B
F
" ... a hybrid CNN model and modify a popular CNN model to predict solar flare occurrence ... "
... ...

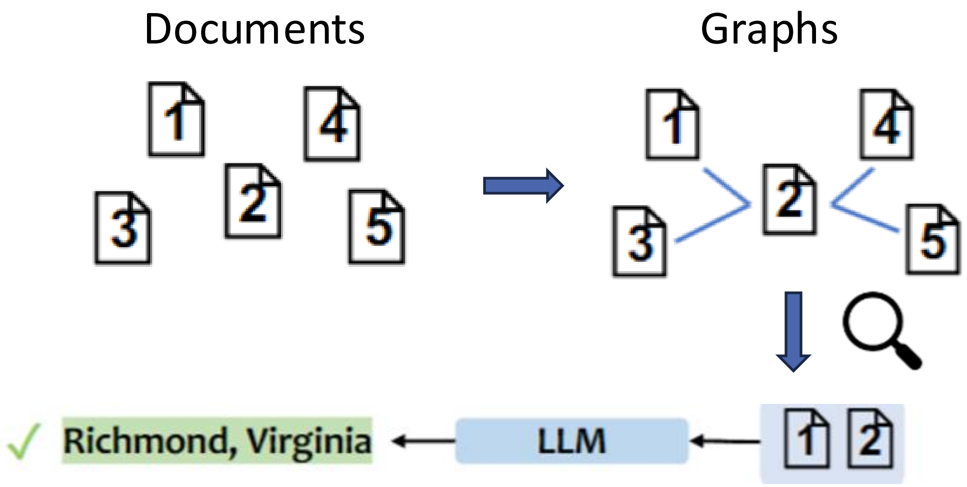
Domain-specific tasks:

- Legal
- Finance
- Medical

…

Graph-enhanced RAG

- **GraphRAG** constructs a graph structure to explicitly model relationships, allowing for more effective and efficient retrieval based on it.



Graph-enhanced RAG

1. Graph Construction

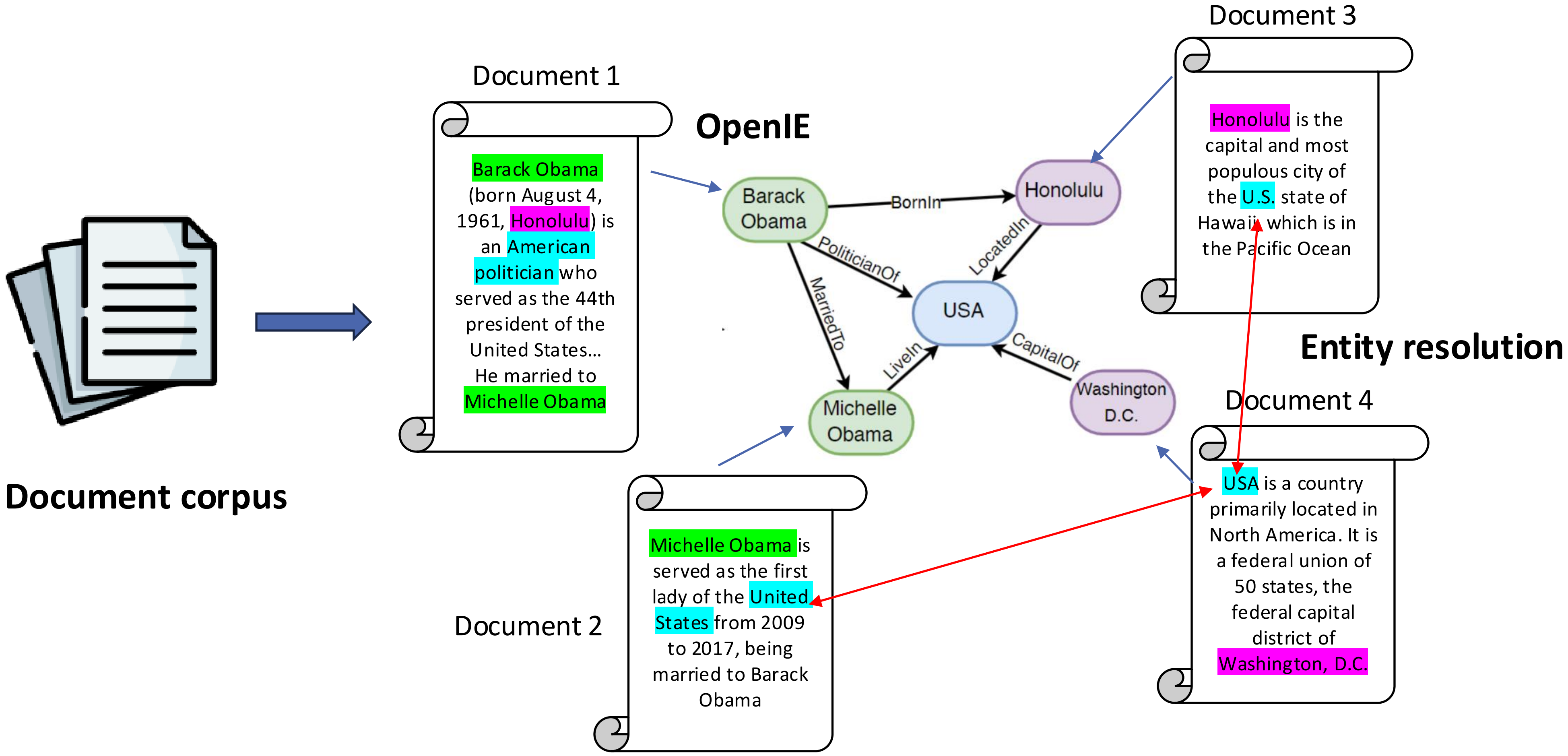
- Hyperlinks
- Reference link
- …

2. Graph-enhanced Retrieval

- Graph search
- GNN
- ….

Knowledge Graph Index

- Knowledge graphs provide a structural index of knowledge across multiple documents are widely used in GraphRAG.
  - KGs can be automatically constructed from documents by LLMs



Document corpus

Document 1

Barack Obama (born August 4, 1961, Honolulu) is an American politician who served as the 44th president of the United States… He married to Michelle Obama

Document 2

Michelle Obama is served as the first lady of the United States from 2009 to 2017, being married to Barack Obama

Document 3

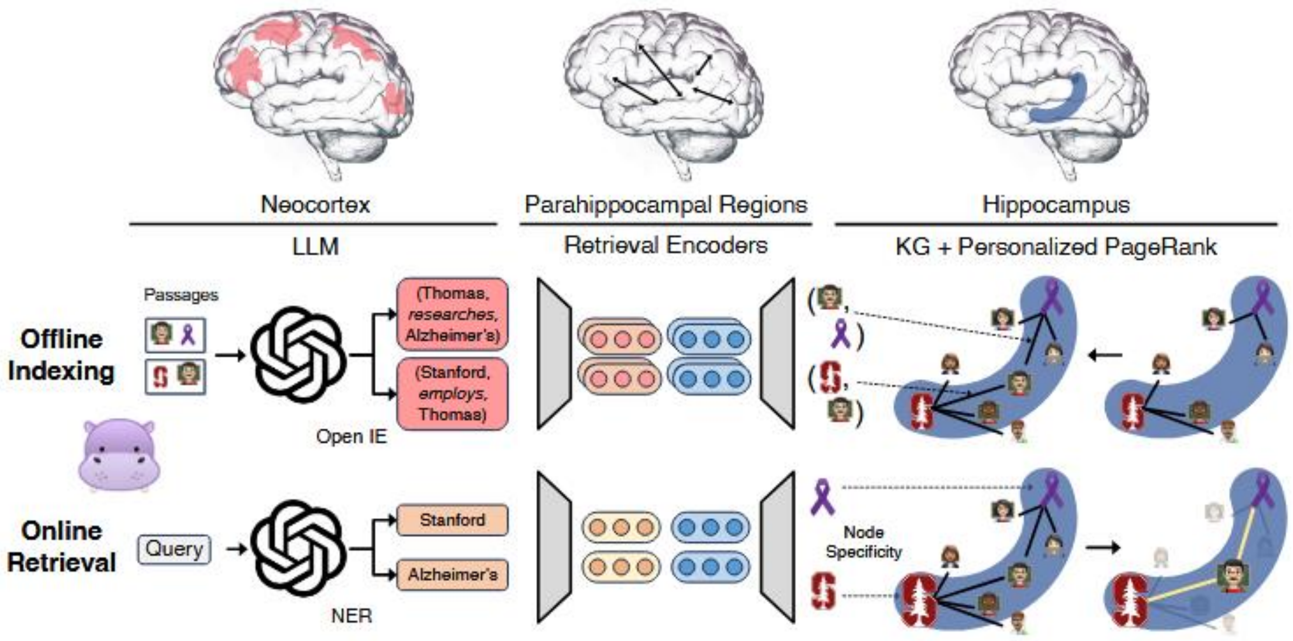
Honolulu is the capital and most populous city of the U.S. state of Hawaii which is in the Pacific Ocean

Document 4

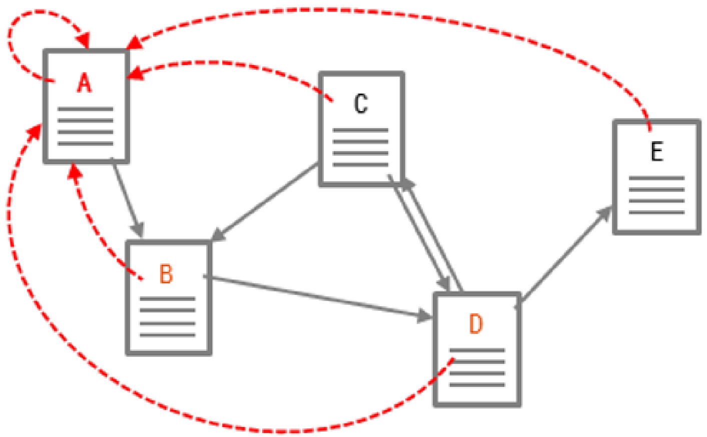
USA is a country primarily located in North America. It is a federal union of 50 states, the federal capital district of Washington, D.C.

HippoRAG (NeurIPS 2024)

- HippoRAG adopts **Personalized PageRank (PPR)** to compute the relevance of documents with the graph structure for retrieval.



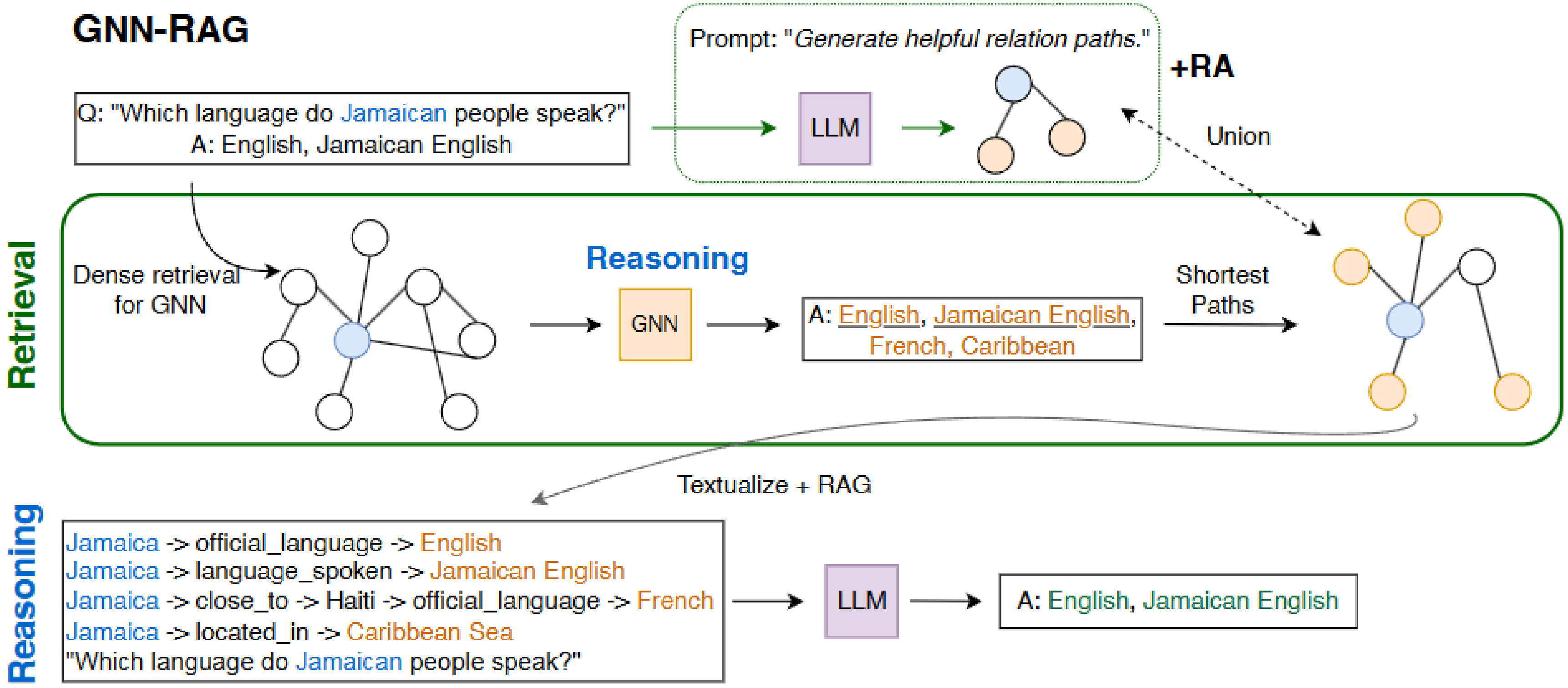
$$\boldsymbol{R}_{i+1} = (1-\alpha)(S^T + w \times d^T)\boldsymbol{R}_i + \alpha \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}$$



Challenge: The graph structure can be noisy and incomplete.

GNN-RAG (ACL 2025)

- **GNNs** have demonstrated impressive performance in GraphRAG due to their powerful graph reasoning ability.

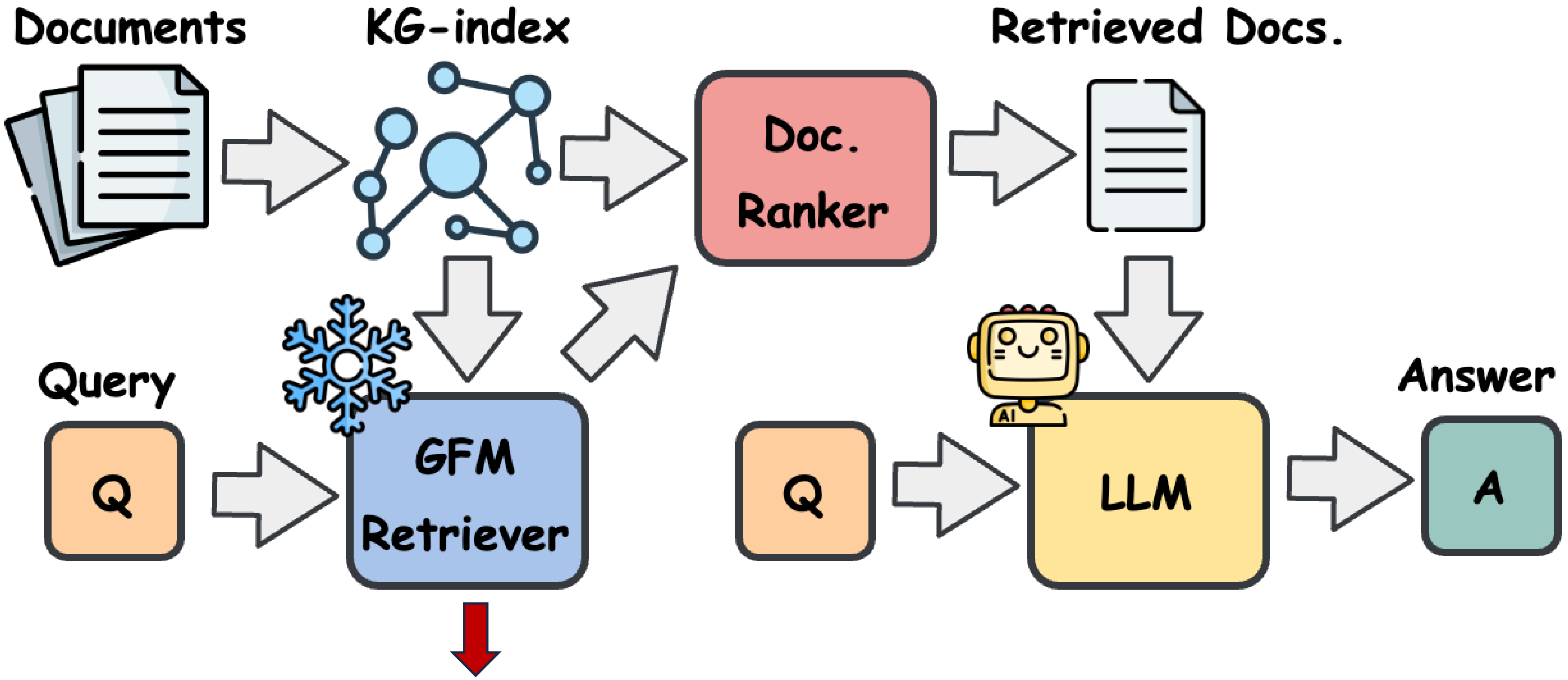

GNN-RAG
Q: "Which language do Jamaican people speak?"
A: English, Jamaican English
Prompt: "Generate helpful relation paths."
LLM
+RA
Union
Retrieval
Dense retrieval for GNN
Reasoning
GNN
A: English, Jamaican English, French, Caribbean
Shortest Paths
Textualize + RAG
Reasoning
LLM
A: English, Jamaican English

Jamaica -> official_language -> English
Jamaica -> language_spoken -> Jamaican English
Jamaica -> close_to -> Haiti -> official_language -> French
Jamaica -> located_in -> Caribbean Sea
"Which language do Jamaican people speak?"

Challenge: GNNs limit in generalizability as they need to be trained from scratch in new datasets.

Graph Foundation Model for Retrieval Augmented Generation

- We propose a novel graph foundation model (GFM), with 8M parameters for retrieval-augmented generation (GFM-RAG).

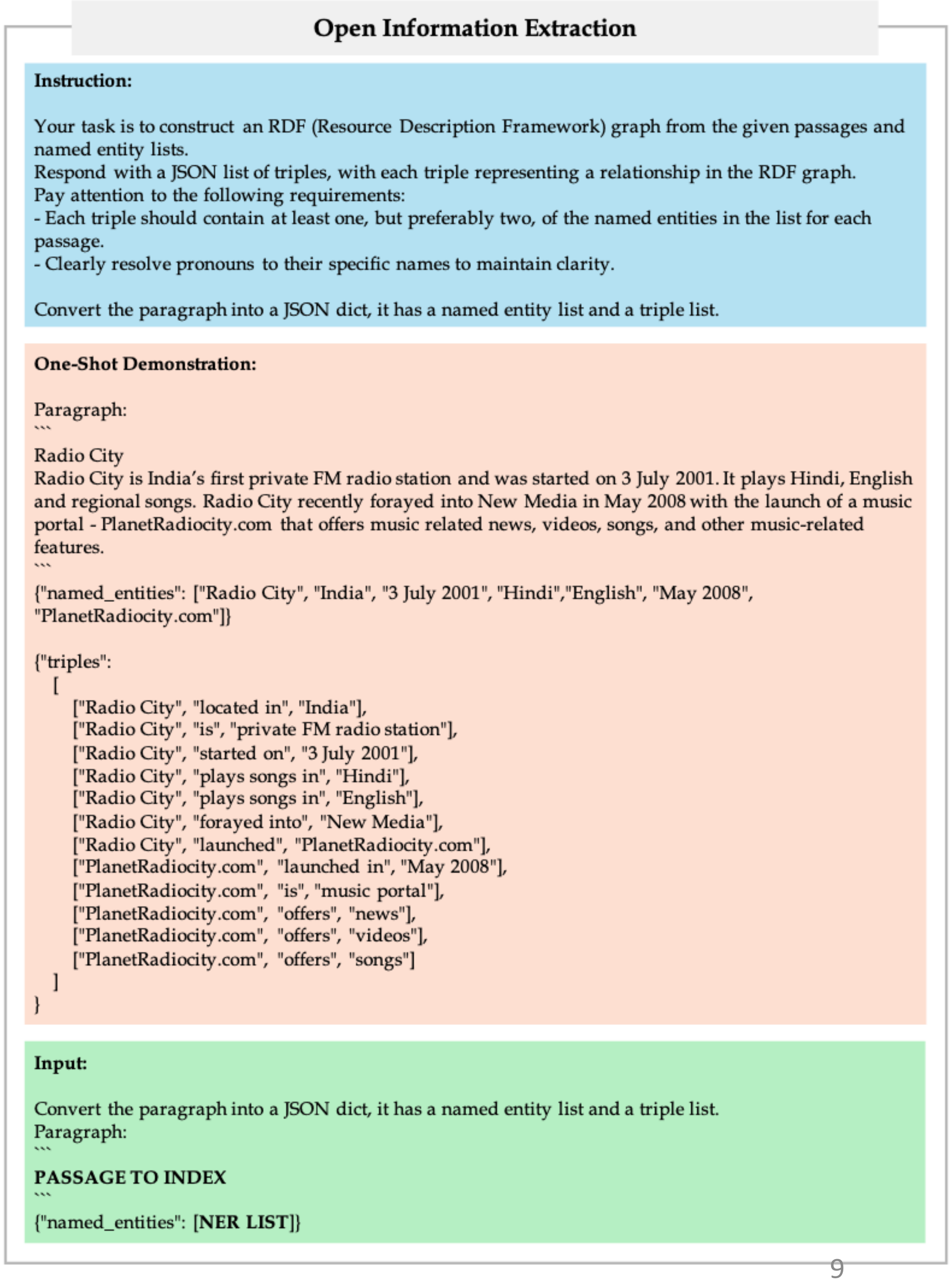


First GFM for RAG that aligns with the neural scaling law!

KG-index Construction

- **OpenIE:** gpt-4o-mini
- **Entity resolution:** colbert
  - Calculate the entities’ embedding similarities and link entities with similar semantics by threshold σ.

$$s = h_{e_1}^T h_{e_2}, s > \sigma$$



Open Information Extraction

Instruction:

Your task is to construct an RDF (Resource Description Framework) graph from the given passages and named entity lists.
Respond with a JSON list of triples, with each triple representing a relationship in the RDF graph.
Pay attention to the following requirements:
- Each triple should contain at least one, but preferably two, of the named entities in the list for each passage.
- Clearly resolve pronouns to their specific names to maintain clarity.

Convert the paragraph into a JSON dict, it has a named entity list and a triple list.

One-Shot Demonstration:

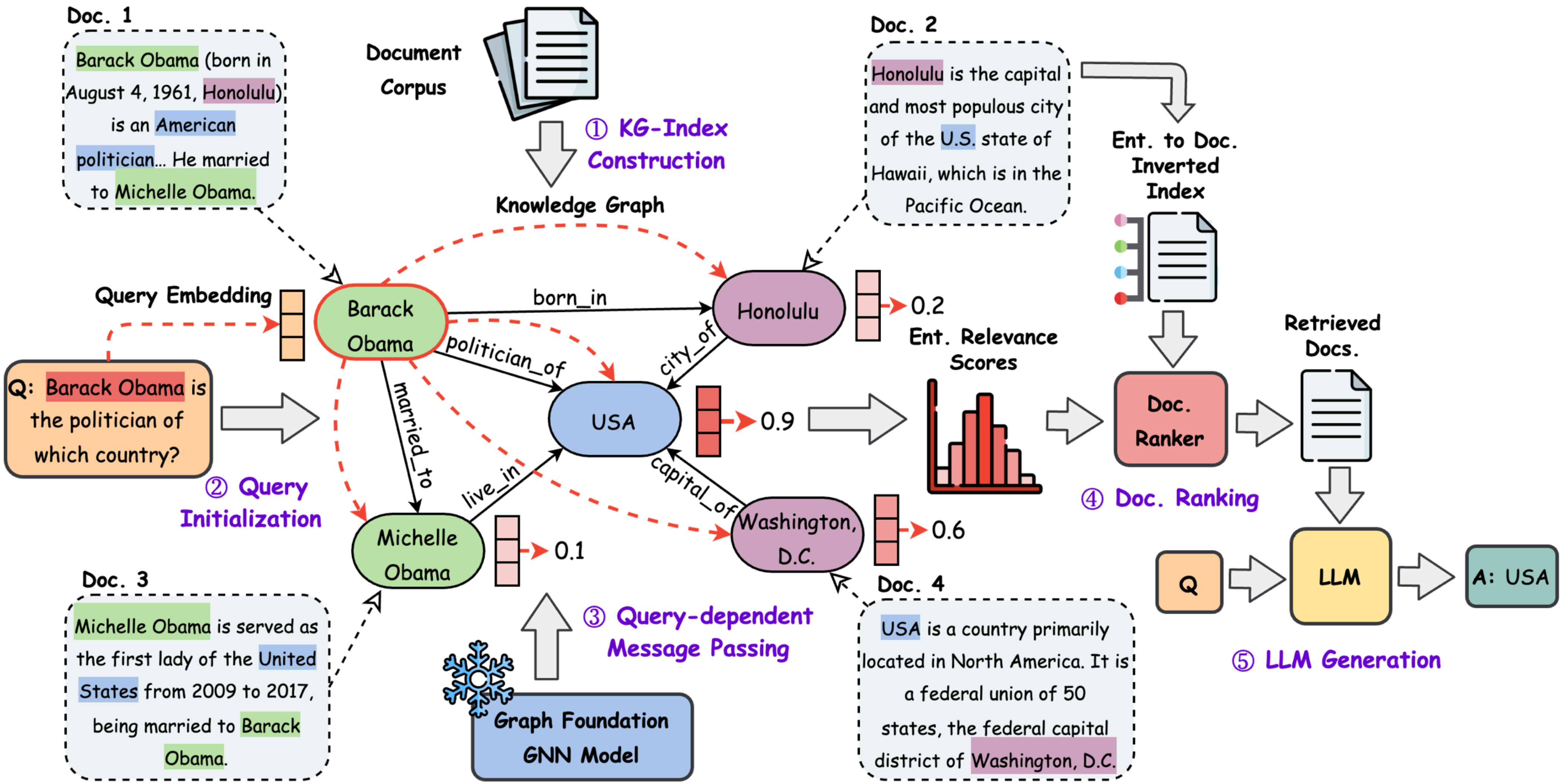
~~~
Paragraph:
```
Radio City
Radio City is India’s first private FM radio station and was started on 3 July 2001. It plays Hindi, English and regional songs. Radio City recently forayed into New Media in May 2008 with the launch of a music portal - PlanetRadiocity.com that offers music related news, videos, songs, and other music-related features.
```
{"named_entities": ["Radio City", "India", "3 July 2001", "Hindi","English", "May 2008", "PlanetRadiocity.com"]}

{"triples":
  [
    ["Radio City", "located in", "India"],
    ["Radio City", "is", "private FM radio station"],
    ["Radio City", "started on", "3 July 2001"],
    ["Radio City", "plays songs in", "Hindi"],
    ["Radio City", "plays songs in", "English"],
    ["Radio City", "forayed into", "New Media"],
    ["Radio City", "launched", "PlanetRadiocity.com"],
    ["PlanetRadiocity.com", "launched in", "May 2008"],
    ["PlanetRadiocity.com", "is", "music portal"],
    ["PlanetRadiocity.com", "offers", "news"],
    ["PlanetRadiocity.com", "offers", "videos"],
    ["PlanetRadiocity.com", "offers", "songs"]
  ]
}
~~~

Input:

~~~
Convert the paragraph into a JSON dict, it has a named entity list and a triple list.
Paragraph:
```
PASSAGE TO INDEX
```
{"named_entities": [NER LIST]}
~~~

Figure 9: Prompt for OpenIE during indexing.

Graph Foundation Model for Retrieval Augmented Generation



Graph Foundation Model for Retrieval Augmented Generation

Doc. 1

Barack Obama (born in August 4, 1961, Honolulu) is an American politician... He married to Michelle Obama.

Doc. 2

Honolulu is the capital and most populous city of the U.S. state of Hawaii, which is in the Pacific Ocean.

Doc. 3

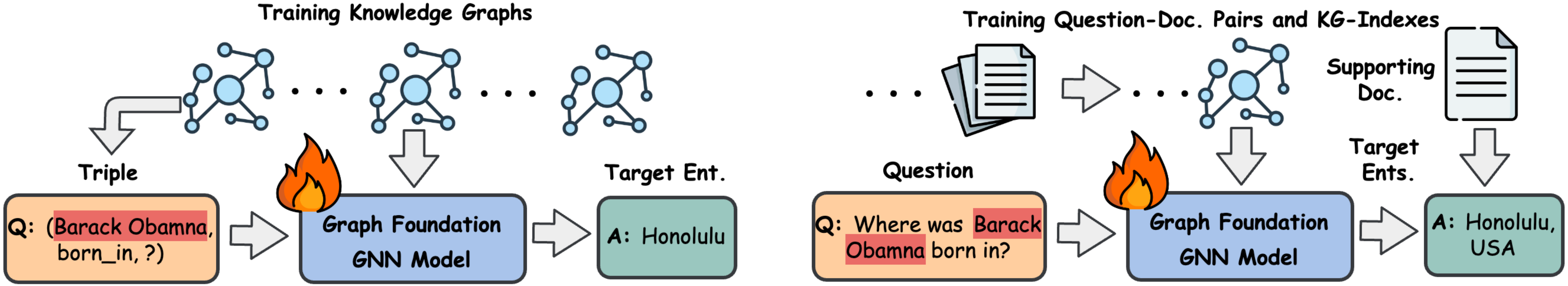
Michelle Obama is served as the first lady of the United States from 2009 to 2017, being married to Barack Obama.

Doc. 4

USA is a country primarily located in North America. It is a federal union of 50 states, the federal capital district of Washington, D.C.

Q: Barack Obama is the politician of which country?

Training Graph Foundation Model



Training Knowledge Graphs

Triple

Q: (Barack Obamna, born_in, ?)

Graph Foundation GNN Model

Target Ent.

A: Honolulu

Training Question-Doc. Pairs and KG-Indexes

Supporting Doc.

Question

Q: Where was Barack Obamna born in?

Graph Foundation GNN Model

Target Ents.

A: Honolulu, USA

Stage 1: Self-supervised KG Completion Pre-training

Stage 2: Supervised Document Retrieval Finetuning

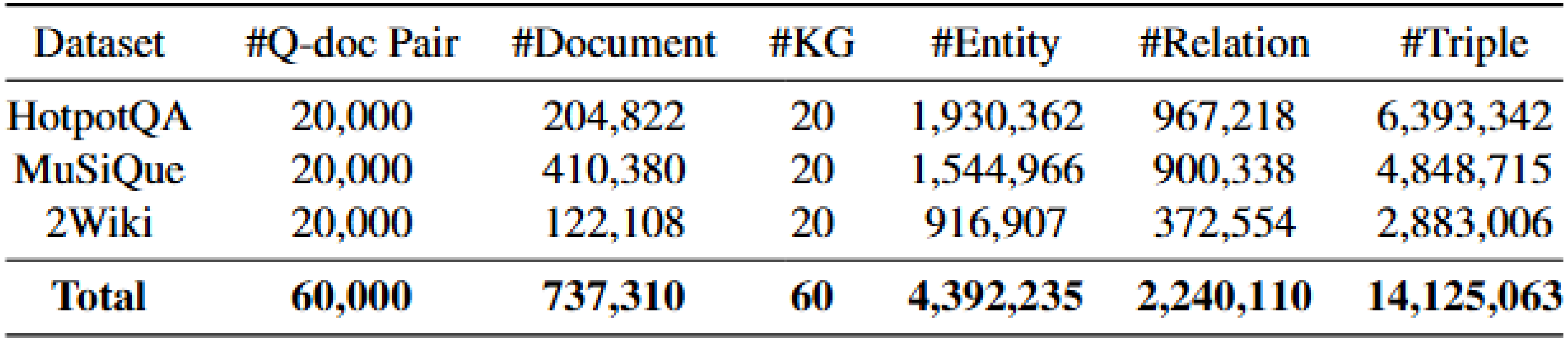
Synthetic query-target pairs

Labeled query-target pairs

Experiments

- **Datasets:**
  - HotpotQA
  - MuSiQue
  - 2Wiki
- **Training:** 8 A100s

Table 1. Statistics of the query-doc pairs and KGs used for training.



Dataset	#Q-doc Pair	#Document	#KG	#Entity	#Relation	#Triple
HotpotQA	20,000	204,822	20	1,930,362	967,218	6,393,342
MuSiQue	20,000	410,380	20	1,544,966	900,338	4,848,715
2Wiki	20,000	122,108	20	916,907	372,554	2,883,006
Total	**60,000**	**737,310**	**60**	**4,392,235**	**2,240,110**	**14,125,063**

Retrieval Performance

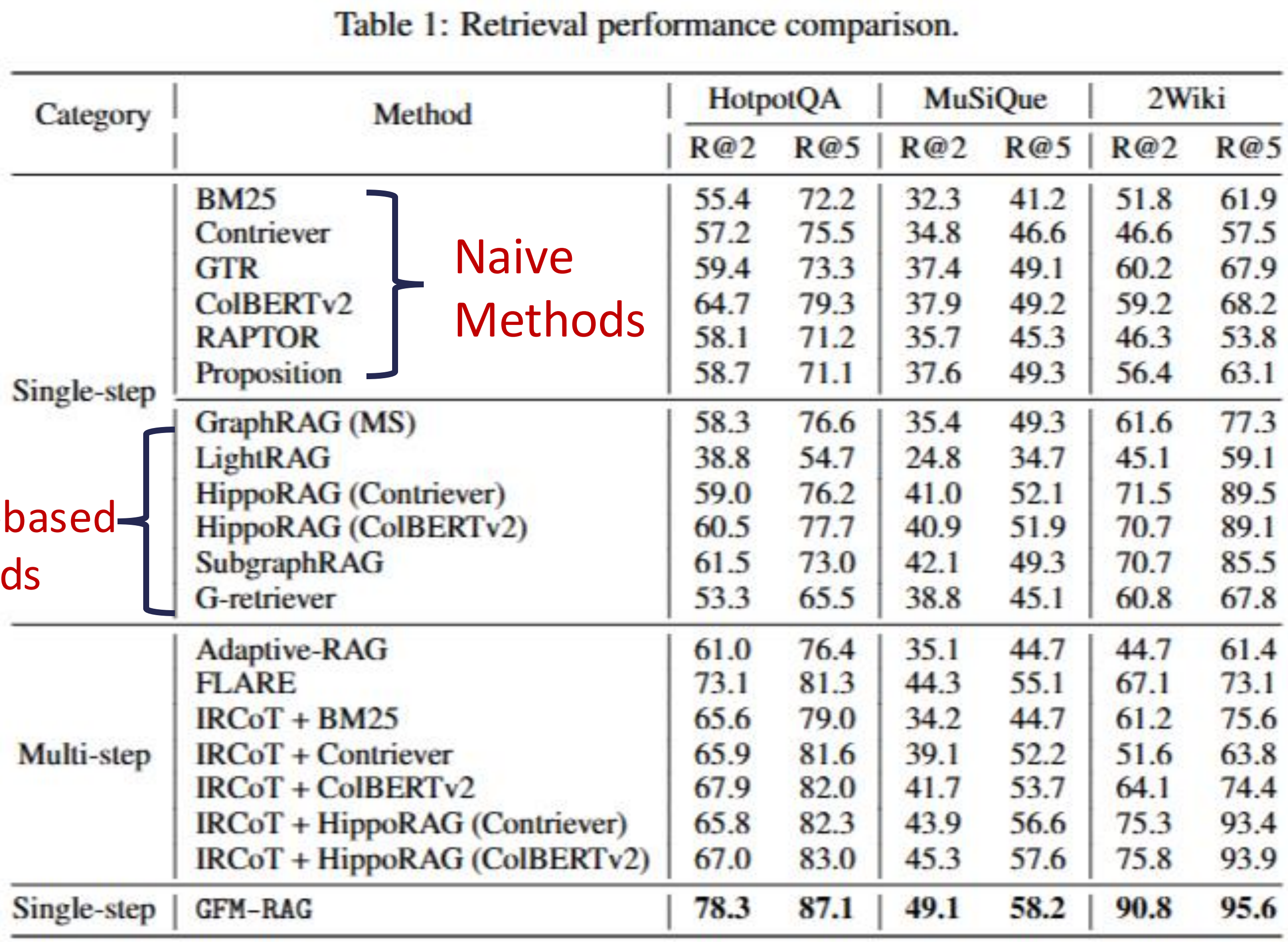


Table 1: Retrieval performance comparison.

Category	Method	HotpotQA		MuSiQue		2Wiki	
		R@2	R@5	R@2	R@5	R@2	R@5
Single-step	BM25	55.4	72.2	32.3	41.2	51.8	61.9
	Contriever	57.2	75.5	34.8	46.6	46.6	57.5
	GTR	59.4	73.3	37.4	49.1	60.2	67.9
	ColBERTv2	64.7	79.3	37.9	49.2	59.2	68.2
	RAPTOR	58.1	71.2	35.7	45.3	46.3	53.8
	Proposition	58.7	71.1	37.6	49.3	56.4	63.1
	GraphRAG (MS)	58.3	76.6	35.4	49.3	61.6	77.3
	LightRAG	38.8	54.7	24.8	34.7	45.1	59.1
	HippoRAG (Contriever)	59.0	76.2	41.0	52.1	71.5	89.5
	HippoRAG (ColBERTv2)	60.5	77.7	40.9	51.9	70.7	89.1
	SubgraphRAG	61.5	73.0	42.1	49.3	70.7	85.5
	G-retriever	53.3	65.5	38.8	45.1	60.8	67.8
Multi-step	Adaptive-RAG	61.0	76.4	35.1	44.7	44.7	61.4
	FLARE	73.1	81.3	44.3	55.1	67.1	73.1
	IRCoT + BM25	65.6	79.0	34.2	44.7	61.2	75.6
	IRCoT + Contriever	65.9	81.6	39.1	52.2	51.6	63.8
	IRCoT + ColBERTv2	67.9	82.0	41.7	53.7	64.1	74.4
	IRCoT + HippoRAG (Contriever)	65.8	82.3	43.9	56.6	75.3	93.4
	IRCoT + HippoRAG (ColBERTv2)	67.0	83.0	45.3	57.6	75.8	93.9
Single-step	GFM-RAG	**78.3**	**87.1**	**49.1**	**58.2**	**90.8**	**95.6**

Naive Methods

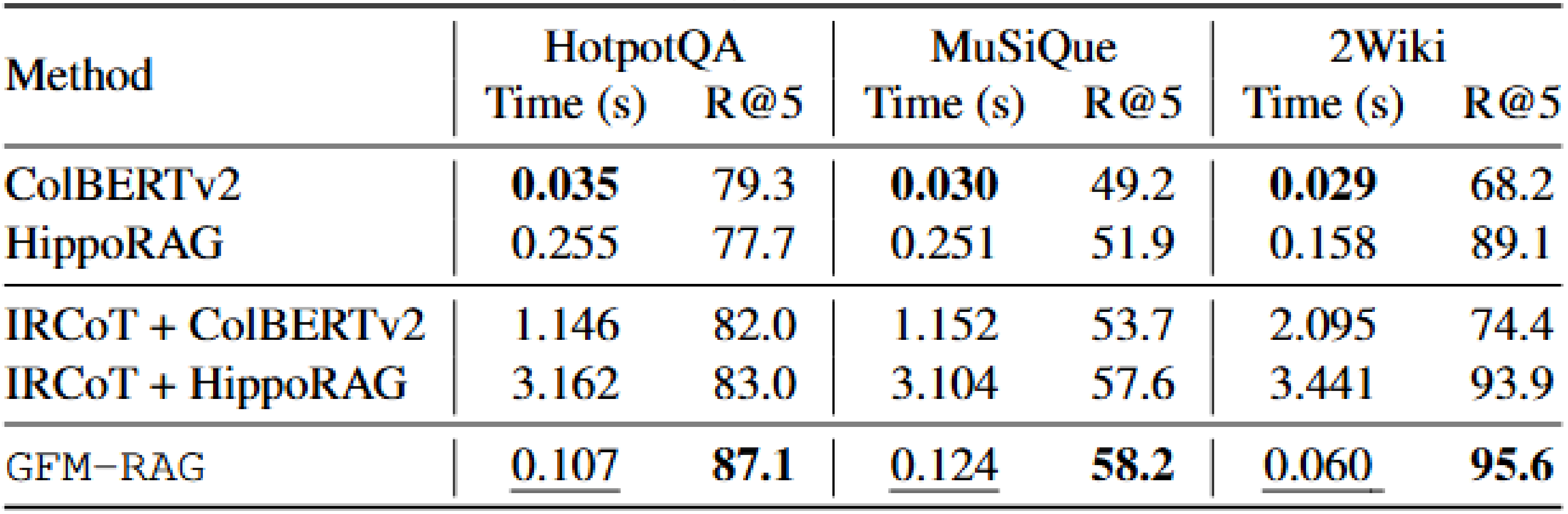
Graph-based Methods

Findings:

- Graph-based method (HippoRAG) > naïve methods.
- Multi-step framework can improve the performance.
- GFM-RAG can effectively conduct the multi-hop reasoning in a single step.

Efficiency

Table 4. Retrieval efficiency and performance comparison.



Method	HotpotQA		MuSiQue		2Wiki	
	Time (s)	R@5	Time (s)	R@5	Time (s)	R@5
ColBERTv2	**0.035**	79.3	**0.030**	49.2	**0.029**	68.2
HippoRAG	0.255	77.7	0.251	51.9	0.158	89.1
IRCoT + ColBERTv2	1.146	82.0	1.152	53.7	2.095	74.4
IRCoT + HippoRAG	3.162	83.0	3.104	57.6	3.441	93.9
GFM-RAG	0.107	**87.1**	0.124	**58.2**	0.060	**95.6**

Findings:

- GFM-RAG achieves a great efficiency in performing multi-step reasoning in a single step.

Generalizability

- Zero-shot transfer to new datasets

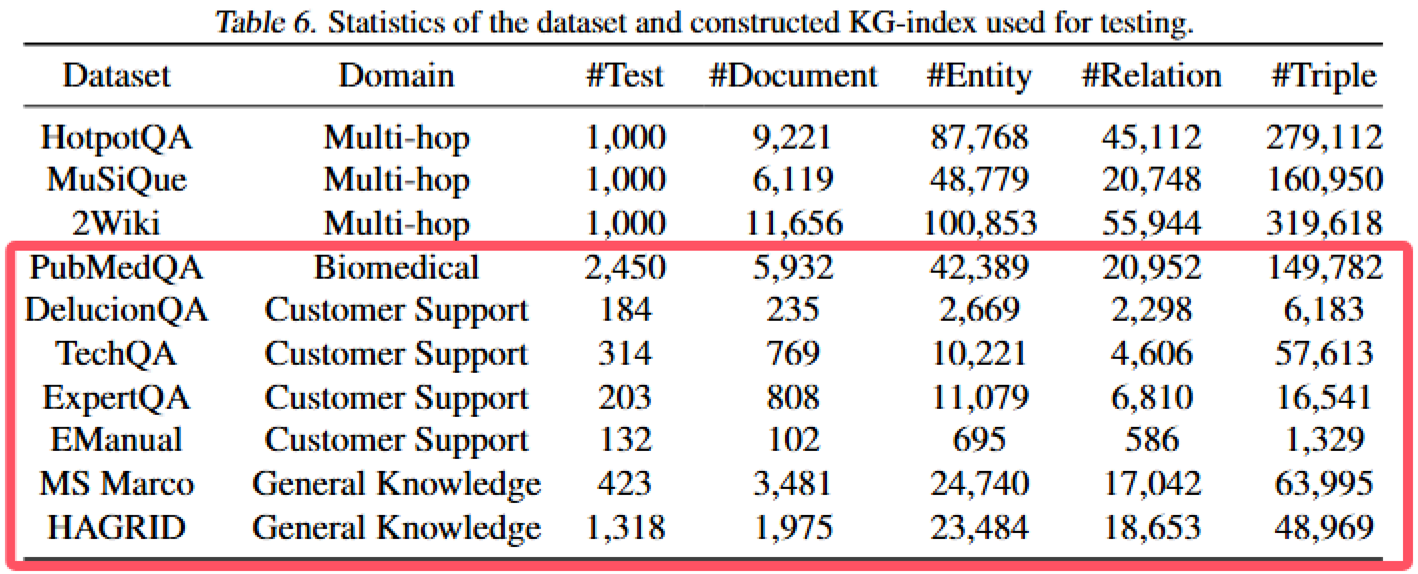
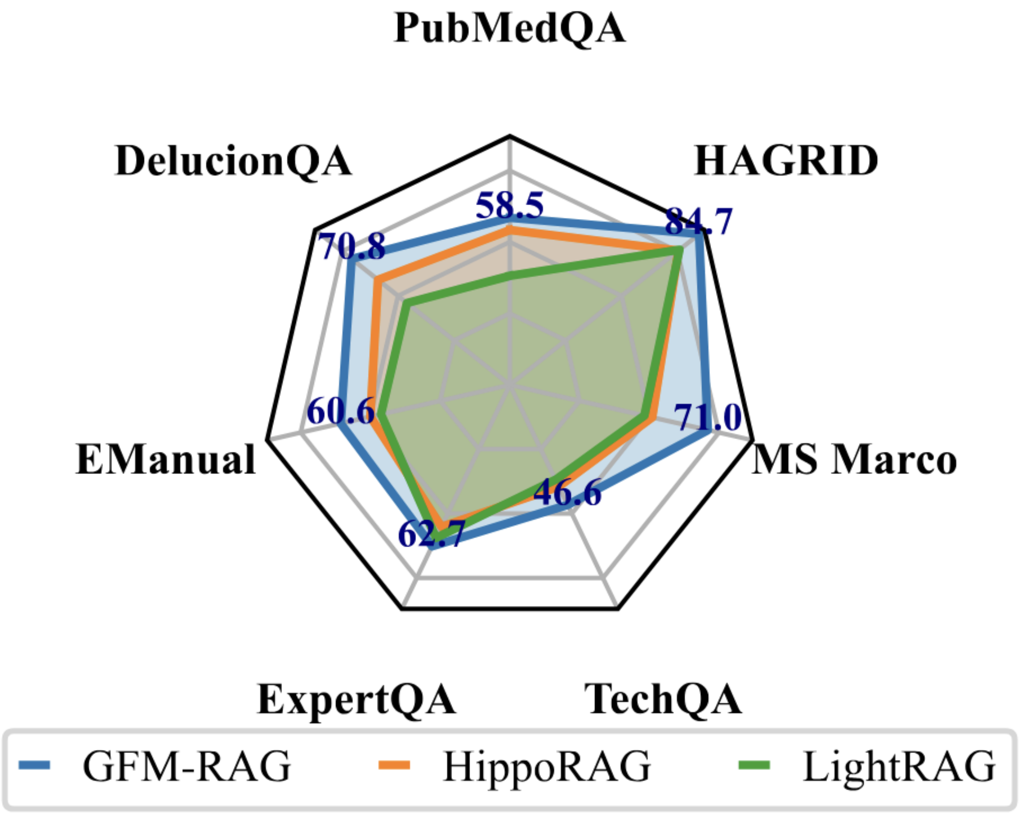


Table 6. Statistics of the dataset and constructed KG-index used for testing.

Dataset	Domain	#Test	#Document	#Entity	#Relation	#Triple
HotpotQA	Multi-hop	1,000	9,221	87,768	45,112	279,112
MuSiQue	Multi-hop	1,000	6,119	48,779	20,748	160,950
2Wiki	Multi-hop	1,000	11,656	100,853	55,944	319,618
PubMedQA	Biomedical	2,450	5,932	42,389	20,952	149,782
DelucionQA	Customer Support	184	235	2,669	2,298	6,183
TechQA	Customer Support	314	769	10,221	4,606	57,613
ExpertQA	Customer Support	203	808	11,079	6,810	16,541
EManual	Customer Support	132	102	695	586	1,329
MS Marco	General Knowledge	423	3,481	24,740	17,042	63,995
HAGRID	General Knowledge	1,318	1,975	23,484	18,653	48,969



Path Interpretations

- GFM can provide path interpretations for its reasoning.

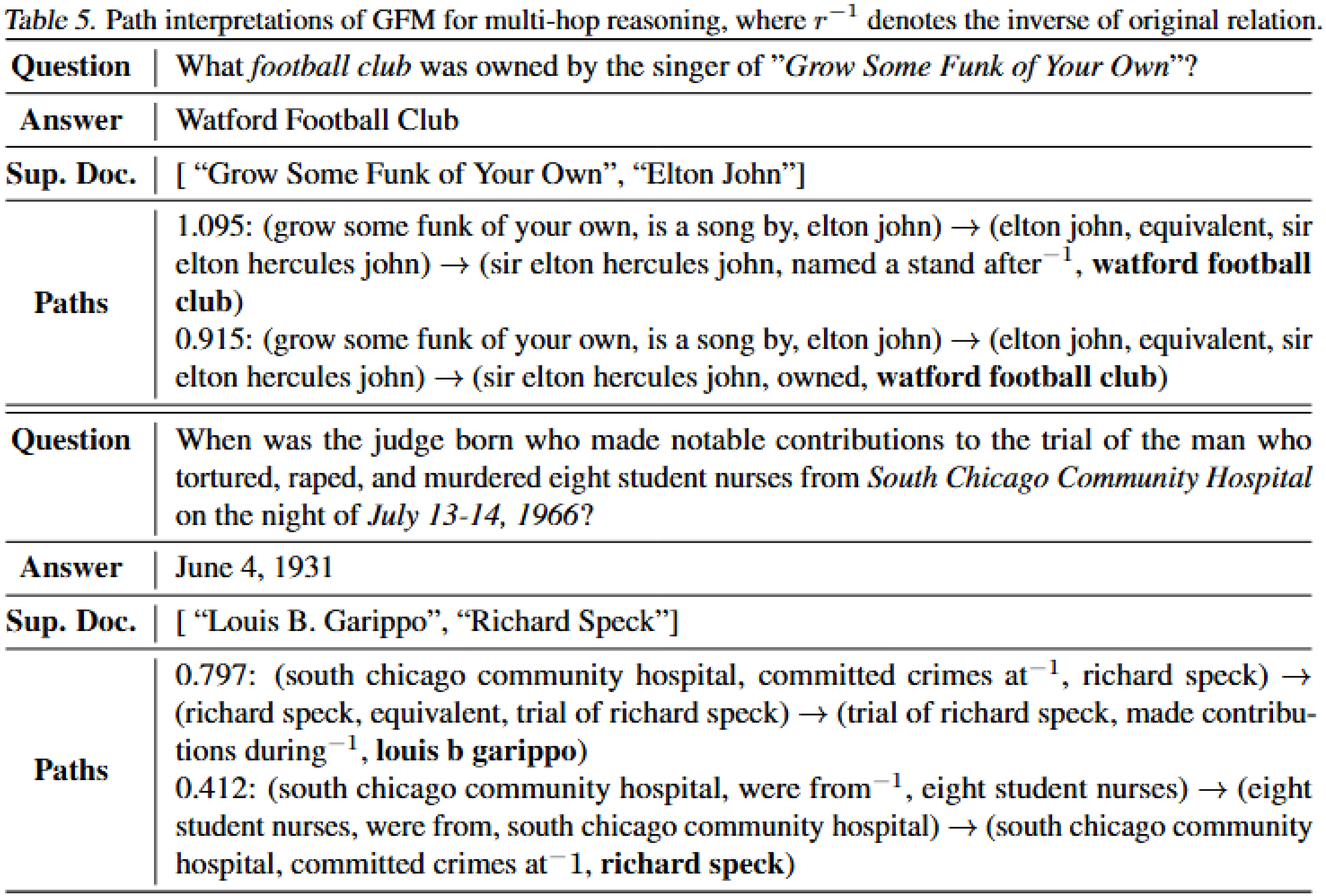


Table 5. Path interpretations of GFM for multi-hop reasoning, where r^{-1} denotes the inverse of original relation.

Question	What *football club* was owned by the singer of "*Grow Some Funk of Your Own*"?
Answer	Watford Football Club
Sup. Doc.	["Grow Some Funk of Your Own", "Elton John"]
Paths	1.095: (grow some funk of your own, is a song by, elton john) → (elton john, equivalent, sir elton hercules john) → (sir elton hercules john, named a stand after^{-1}, **watford football club**) 0.915: (grow some funk of your own, is a song by, elton john) → (elton john, equivalent, sir elton hercules john) → (sir elton hercules john, owned, **watford football club**)
Question	When was the judge born who made notable contributions to the trial of the man who tortured, raped, and murdered eight student nurses from *South Chicago Community Hospital* on the night of *July 13-14, 1966*?
Answer	June 4, 1931
Sup. Doc.	["Louis B. Garippo", "Richard Speck"]
Paths	0.797: (south chicago community hospital, committed crimes at^{-1}, richard speck) → (richard speck, equivalent, trial of richard speck) → (trial of richard speck, made contributions during^{-1}, **louis b garippo**) 0.412: (south chicago community hospital, were from^{-1}, eight student nurses) → (eight student nurses, were from, south chicago community hospital) → (south chicago community hospital, committed crimes at^{-}1, **richard speck**)

The path's importance to the final prediction can be quantified by the partial derivative of the prediction score with respect to the triples at each layer.

$$s_1, s_2, \ldots, s_L = \text{top-}k \frac{\partial p_e(q)}{\partial s_*}.$$

Neural scaling law

- Performance of the GFM scales with the data and parameters.

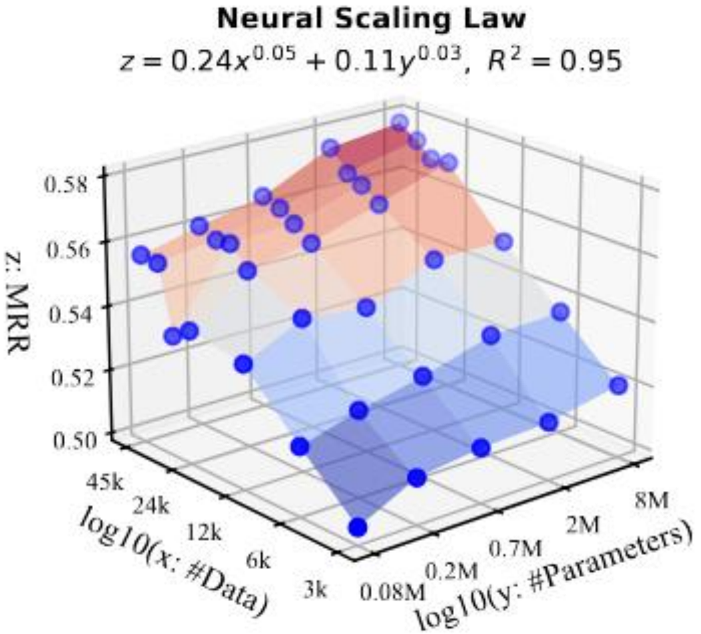


Figure 4. Neural scaling law of GFM−RAG.

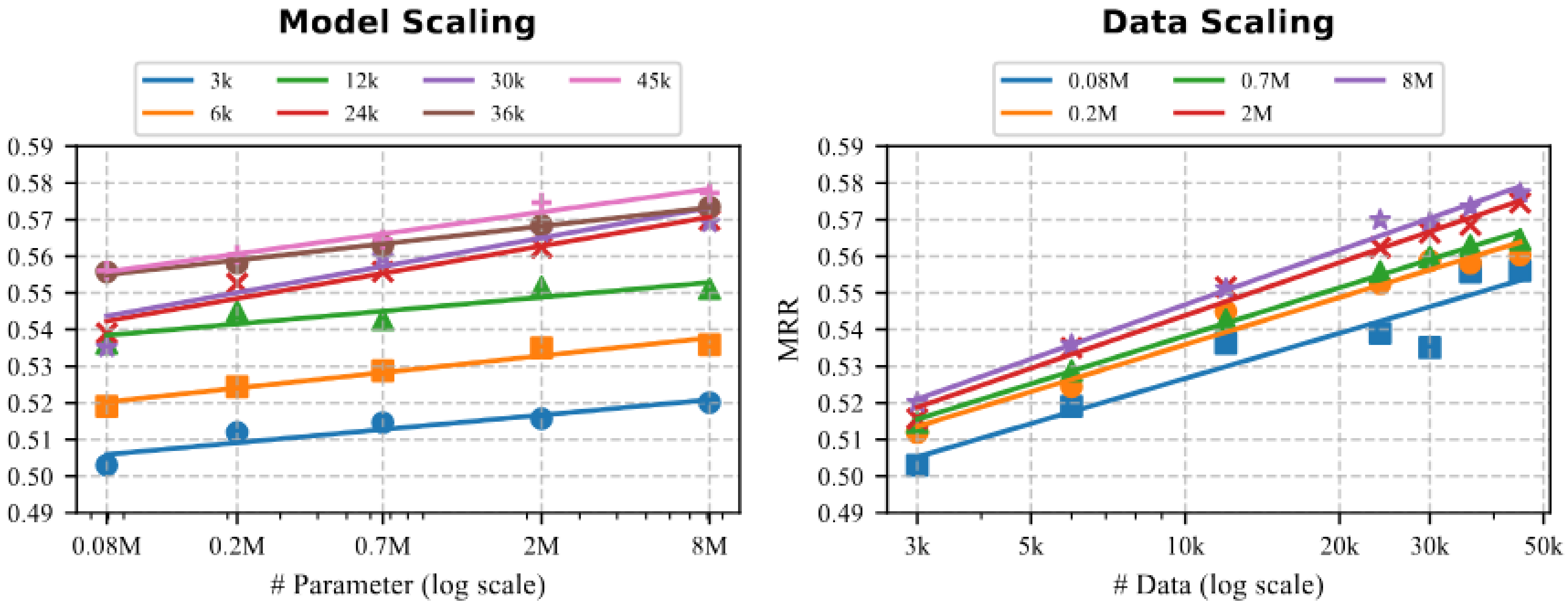
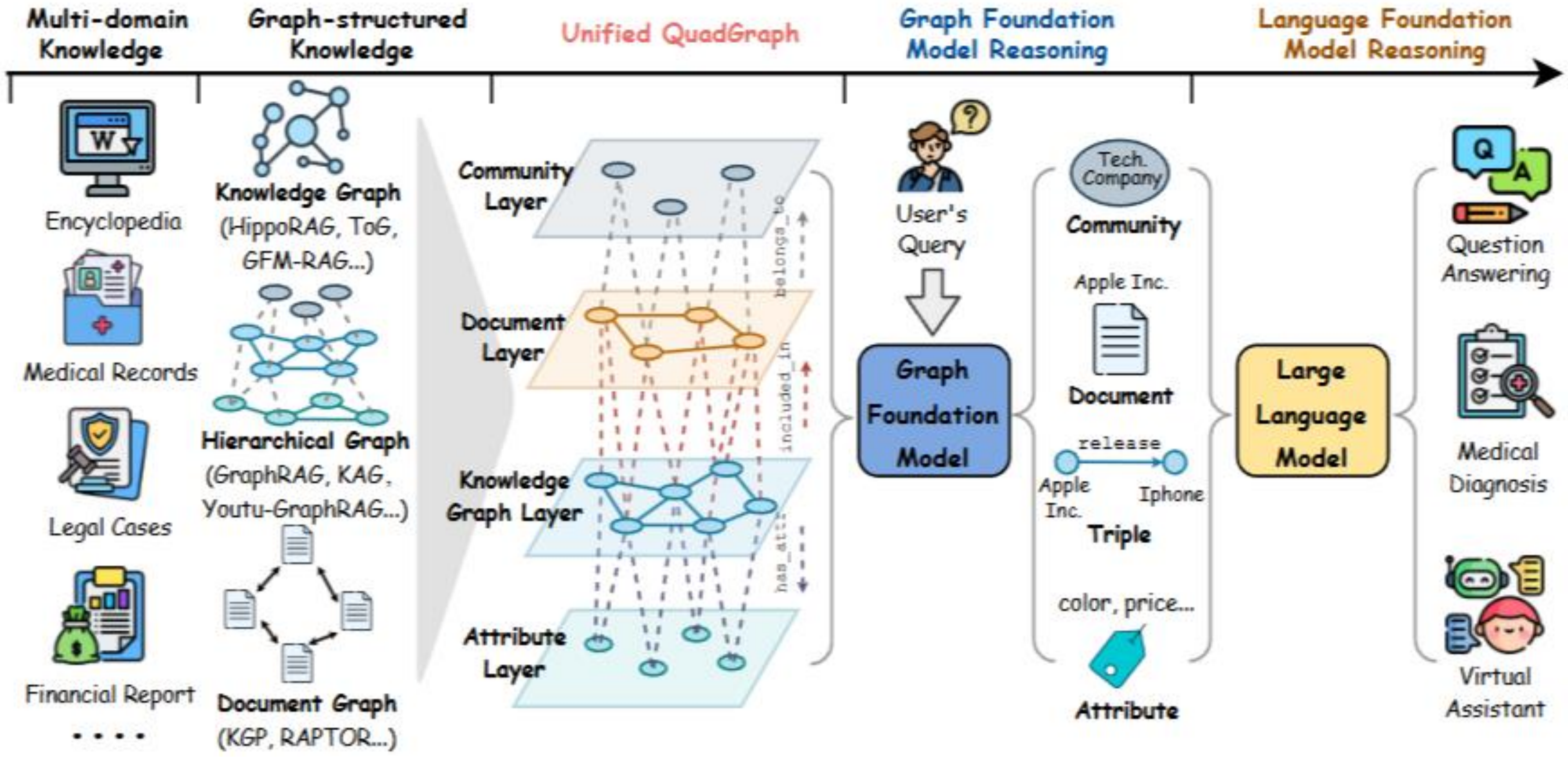


Figure 5. The illustration of the model and data scaling law of GFM−RAG.

Future works – G-reasoner

- **G-reasoner: Foundation Models for Unified Reasoning over Graph-structured Knowledge**



Paper

Thanks for your listening!

Paper

Code