

A Comprehensive Graph Framework for Question Answering with Mode-Seeking Preference Alignment

Quanwei Tang¹, Sophia Yat Mei Lee², Junshuang Wu³, Dong Zhang^{1*}, Shoushan Li¹, Erik Cambria⁴ and Guodong Zhou¹

¹School of Computer Science & Technology, NLP Lab, Soochow University, China
²Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
³Beijing Jinhang Research Institute of Computing and Communication, China
⁴College of Computing and Data Science, Nanyang Technological University, Singapore
dzhang@suda.edu.cn

Abstract

Recent advancements in retrieval-augmented generation (RAG) have enhanced large language models in question answering by integrating external knowledge. However, challenges persist in achieving global understanding and aligning responses with human ethical and quality preferences. To address these issues, we propose GraphMPA, a comprehensive graph-based framework with mode-seeking preference alignment. Our approach constructs a hierarchical document graph using a general similarity measurement, mimicking human cognitive processes for information understanding and synthesis. Additionally, we introduce mode-seeking preference optimization to better align model outputs with human preferences through probability-matching constraints. Extensive experiments on six datasets demonstrate the effectiveness of our GraphMPA.

1 Introduction

Retrieval-augmented generation (RAG) with large language models (LLMs) has recently emerged as a promising approach in question-answering (QA) (Zhao et al., 2024; Gao et al., 2024). This is mainly due to its ability to retrieve external documents, thus increasing the knowledge of the model. However, despite its advances, existing studies still face the following challenges at both input and output levels.

From the representation and understanding of **input external documents**, previous works (Barnett et al., 2024; Asai et al., 2024; Xu et al., 2024; Lewis et al., 2021; Liu et al., 2024) often struggle with tasks that require comprehensive, global understanding, as they fail to provide a unified and holistic view of the relevant external knowledge. For instance, as shown in Figure 1, if we directly retrieve effective information across all input documents for an overarching query like "What about

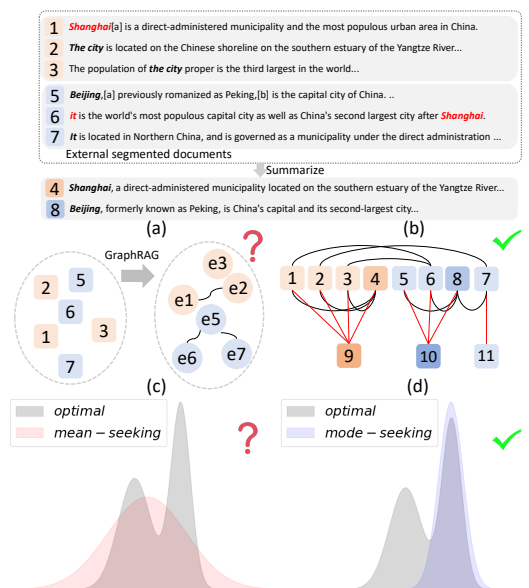


Figure 1: (a) Prior entity-based graph and (b) Our hierarchical graph with community summarization. (c) Prior DPO with large-scale LLMs generated data and (d) Our MS with small-scale LLMs synthetic data.

Shanghai?", the potential answer may refer to sentence 6, where the pronoun "it" is interpreted as referring to "Shanghai" (it actually refers to Beijing.) due to the presence of the query words at the end of this sentence. Only very recently, Guo et al. (2024) and Edge et al. (2024) introduced a global graph-based RAG strategy to alleviate this issue, but it heavily relies on large-scale LLMs (e.g., GPT-4) to extract entity pairs and their corresponding relations from each sentence as Figure 1(a). This not only consumes significant resources but may also overlook sentences without entities or relations like the sentence 3 in Figure 1. Therefore, we argue for the use of smaller, more efficient LLMs and the construction of sentence relevance using a more general and comprehensive measurement as Figure 1(b).

*Corresponding Author: Dong Zhang

From the quality of **output answers**, ensuring that LLMs generate responses aligned with human standards and preferences remains a significant statistical and practical challenge in QA (Oneto et al., 2016; Cambria et al., 2013), as it involves modeling complex patterns of human reasoning, contextual understanding, and nuanced language use. Although approaches like DPO (Rafailov et al., 2024) have been proposed to mitigate this issue, they lead to a compromised approximation of optimal distribution (Lin et al., 2021b; Wu et al., 2024) based on large-scale LLMs (GPT-4) generated corpus, such as a *mean-seeking* policy that places large mass to the mean of different modes (Chan et al., 2022) as illustrated in Figure 1(c). However, it is insufficient since we need to estimate the exact and primary mode of the target distribution (Ji et al., 2024). We argue that a *mode-seeking* strategy based on an auto-constructed corpus with small-scale LLMs, can more effectively capture the alignment with human preferences as shown in Figure 1(d).

To address these challenges, in this paper, we propose a comprehensive hierarchical graph framework with mode-seeking preference alignment (**GraphMPA**) based on RAG for QA. Specifically, we first demolish external knowledge into a hierarchical graph using a simple but comprehensive similarity measurement, mimicking human cognitive processes in organizing and synthesizing information through abstractive summarization from low-level to high-level (Sweller, 1988; Chandler and Sweller, 1991). Next, we retrieve the top-K small documents to construct a human-preferred dataset by data synthesis for real preference alignment. Finally, to achieve better mode-seeking, we introduce mode-seeking preference optimization to align the model with human preferences. In general, our contributions are summarized as follows:

- 1) We propose a hierarchical graph with community summarization based on a general similarity measurement, improving comprehensive sentence relevance and global understanding of external candidate knowledge for QA tasks.

- 2) We introduce a mode-seeking preference optimization strategy by applying probability-matching constraints between the parametrized policy and the optimal policy without relying on very large-scale LLMs to generate human-preferred data.

- 3) We carry out extensive experiments and detailed analysis on six representative datasets, demonstrating the effectiveness of our proposed GraphMPA.

2 Related Work

LLMs have achieved significant success in a wide range of QA tasks. However, their notable challenge is the limited access to specialized or up-to-date knowledge, which can lead to outdated or incomplete responses in domain-specific tasks. To address these issues, RAG has been proposed as a solution (Gao et al., 2024). Our work is mainly related to RAG and human preference alignment.

Retrieval-Augmented Generation. RAG enhances LLMs with external retrieval, making it ideal for knowledge-intensive tasks (Huang et al., 2025). Traditional approaches of QA (Tang et al., 2024) normally face the "lost in the middle" problem (Liu et al., 2023), and are expensive and slow. RAG can alleviate these issues by retrieving relevant documents from external sources and incorporating this information into the generation process. However, determining when and which documents to retrieve and global understanding issues remain challenging (Gao et al., 2024). Therefore, Edge et al. (2024) propose a graph-based method for query-focused summarization by extracting an entity knowledge graph with very large-scale LLMs, like GPT-4. Meanwhile, Sarthi et al. (2024) suggest a recursive, hierarchical summarization approach by a tree structure, but it ignores the internal relations among nodes in each layer.

Unlike the above studies, we design a comprehensive graph structure with simple and fine-grained document relations to incorporate effective knowledge, which completely abandons large-scale LLMs as entity-relation extractors.

Human Preference Alignment (HPA). Recently, Direct Preference Optimization (DPO) (Rafailov et al., 2024) has emerged as a promising alternative to the previously popular reinforcement learning from human feedback (RLHF) (Christiano et al., 2023; Ouyang et al., 2022). It simplifies the training pipeline, eliminating the need for separate reward models and policy optimization (Rafailov et al., 2024). However, this often results in a mean-seeking approximation that compromises the optimal solution and relies on GPT-4 generated data for optimization. Consequently, the strategy π_θ tends to cover all modes of the target strategy π^* , rather than concentrating on the most important modes. As a result, the generated texts or behaviors may lack clear direction or focus, failing to effectively capture the key features of human preferences.

Unlike previous works, to avoid the compromised mean-seeking approximation of the optimal solution in standard DPO, we introduce the mode-seeking (MS) loss inspired by (Ji et al., 2024) and design an intuitive auto-constructed dataset with small-scale LLMs for training.

3 Methodology

Motivation. To systematically understand the traditional fragmented pieces of knowledge, we draw inspiration from human cognitive processes in handling and organizing information, proposing to summarize closed small documents and organize them into a *hierarchical graph*. This also enhances both the low-level and the high-level comprehension of external knowledge. Then, we simulate the human thinking process to derive the final answer through structured reasoning. This method leverages the *mode-seeking preference optimization* aiming to better fit the optimal pattern (mode) rather than the overall expectation.

Task Formulation. In RAG, the objective is to answer a natural language question Q by querying a set of segmented documents D . Each document d is typically derived from external text. The question and the retrieved documents are then passed into an LLM \mathcal{M} to generate the answer A . During the retrieval process, an embedding model EMBED is used to transform the text input into dense vector representations.

3.1 Summarization-based Hierarchical Graph

We expect to capture both low-level and high-level information in the input ultra-long text, so we intersperse summarization strategies to construct a hierarchical graph network. This summarization allows retrieval augmentation of context at different scales. Figure 2 shows the overall iterative process of building a graph, including graph initialization, community detection, and hierarchical structuring, which collectively enhance the organization and representation of retrieved information.

Document Splitting and Relations. The entire input candidate text to be retrieved is first divided into large documents (e.g., the whole of sentences 1, 2 and 3 in Figure 1), each of which is then summarized to produce a summary document denoted as $S_D = \{D_1, \dots, D_l\}$ with LLMs. These large documents are recursively subdivided into smaller documents $S'_D = \{D_k, \dots, D_n\}$ (e.g., the sentence 1 or 2 in Figure 1) with $\text{TextSplit}(D_{large}, small)$.

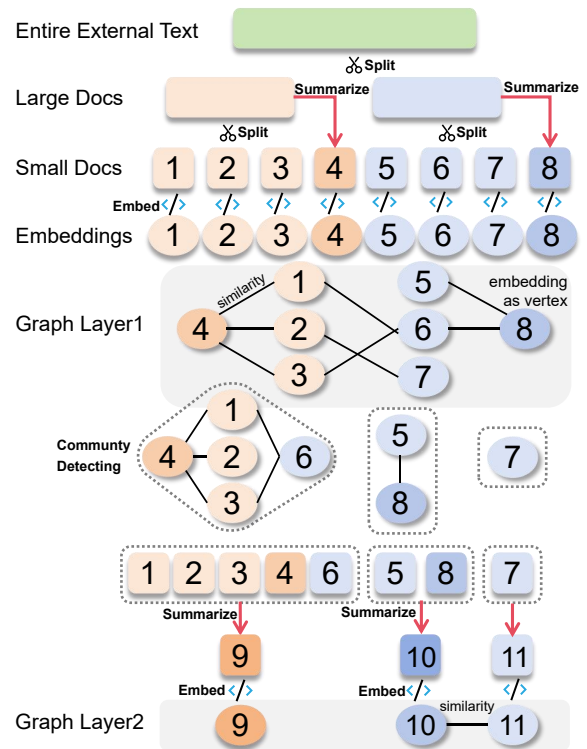


Figure 2: The process of building the hierarchical graph with community summarization.

For each small document D_i , the document embedding e_{D_i} is computed using an embedding model EMBED, like BGE-M3 (Chen et al., 2024; Alwaneen et al., 2022), which transforms the textual content of the document into a dense vector representation. The similarity between the embeddings of documents is then computed by cosine similarity as,

$$\text{sim}(D_i, D_j) = \frac{e_{D_i} \cdot e_{D_j}}{\|e_{D_i}\| \|e_{D_j}\|} \quad (1)$$

To incorporate all proper information, we combine all summary documents S_D and small documents S'_D as $\mathbf{D} = \{D_1, \dots, D_n\}$. Then, a document similarity matrix M is constructed by calculating the pairwise relations among all documents' embeddings.

A Layer of Graph Building. For the convenience of graph calculation, we remove a small number of edges with weak correlation by a threshold τ according to the similarity matrix M . To this end, we construct a layer of the graph, defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:

Vertexes \mathcal{V} : $v \in \mathcal{V}$ represents a small document or summarization document D from large document with its embedding as input. Formally,

$$v = e_D = \text{EMBED}(D) \quad (2)$$

Edges \mathcal{E} : $e \in \mathcal{E}$ represents a similarity between two small documents through embeddings. The weight of edge e is defined:

$$w_{ij} = \begin{cases} s_{ij} & , \quad \text{sim}(D_i, D_j) \geq \tau \\ 0 & , \quad \text{else} \end{cases} \quad (3)$$

This layer of the graph is then added to the set of *layers*.

Community Detection. To better understand complex and lengthy candidate text, we construct a hierarchical graph. We cluster nodes in the current layer to organize segments into cohesive groups, referred to as community detection (CD). These groups are then summarized into a new node for the next layer.

To achieve this, we apply the Leiden algorithm (Traag et al., 2019) to each graph layer, extracting communities. The algorithm ensures internal connectivity and finds high-quality partitions efficiently, making it ideal for large-scale, complex networks. This step clusters related contexts, aiding the retrieval process. Formally,

$$C = \text{CommDetect}(G_i) \quad (4)$$

For each detected community, a community summary S_C is generated, forming a new document D' . This document encapsulates the key aspects of the community’s content, and a new graph G_{i+1} is created based on these summaries. The graph is added to the set of *layers*, and the embedding calculation for each graph is repeated until the desired depth L is reached.

Iterative Hierarchy Building. At each iteration, document embeddings are recalculated, and document similarities are updated. This recursive process refines the system’s understanding of the document space, improving retrieval and performance in tasks like query answering. The whole process can refer to Algorithm 1. The argument *large* denotes the length of large chunks (an integer), *small* represents small documents length (an integer), *LLM.summary* denotes the abstractive summarization by LLMs, and k indicates the number of top k nodes retrieved for related node searching.

3.2 Generating with Preference Alignment

In this section, we introduce a preference alignment approach with a novel optimization strategy that simultaneously emphasizes intermediate reasoning steps and the final answer, enhancing the

Algorithm 1 Build Graph Algorithm

```

1: function BUILD GRAPH(text ▷ Document,
   large, small,           ▷ Output length
   n_layers,               ▷ Depth  $L$ 
    $\tau$                      ▷ Threshold )
2:  $D_{large} \leftarrow \text{TextSplit}(\textit{text}, \textit{large})$ 
3:  $S_D \leftarrow \text{LLM.summary}(D_{large}, \textit{small})$ 
4:  $S'_D \leftarrow \text{TextSplit}(D_{large}, \textit{small})$ 
5:  $D \leftarrow S_D \cup S'_D$ 
6:  $\textit{layers} \leftarrow []$ 
7: while  $n\_layers > 0$  do
8:    $e_D \leftarrow \text{EMBED}(D)$ 
9:    $M \leftarrow \text{sim}(e_D, e_D^T)$ 
10:   $\textit{sims} \leftarrow \text{sort}(M, \textit{axis} \leftarrow 0, \textit{reverse})$ 
11:   $\mathcal{V} \leftarrow \text{arg sort}(M, \textit{axis} \leftarrow 0, \textit{reverse})$ 
12:   $\mathcal{E} \leftarrow []$ 
13:  for  $u \leftarrow 0$  to  $\text{len}(\mathcal{V})$  do
14:    for  $v$  in  $V[u]$  do
15:       $w \leftarrow \textit{sims}[u][v]$ 
16:      if  $w \geq \tau$  then
17:         $E.append((u, v, w))$ 
18:      end if
19:    end for
20:  end for
21:   $\mathcal{G} \leftarrow (\mathcal{V}, \mathcal{E})$ 
22:   $\textit{layers}.append(\mathcal{G})$ 
23:   $n\_layers \leftarrow n\_layers - 1$ 
   ▷ Prepare for the next layer
24:   $C \leftarrow \text{CommDetect}(\mathcal{G})$ 
25:   $D \leftarrow \text{LLM.summary}(C, \textit{small})$ 
26: end while
27: return  $\textit{layers}$ 
28: end function

```

model’s understanding of the reasoning process and enabling more reliable, coherent, and context-aware outputs. Specifically, we first retrieve the most query-relevant small documents. Then, based on these informative documents, we build the contrastive dataset with the **human-preferred answer** with a reasoning process and the **less preferred answer** without any reasoning process. Finally, we propose to leverage the mode-seeking loss to optimize the model more satisfying against traditional DPO.

Retrieval with Semantic Ranking. After understanding both low- and high-level messages of the input long candidate text, we select the most relevant information for the specific query by retrieving top- k small documents on semantic measurement.

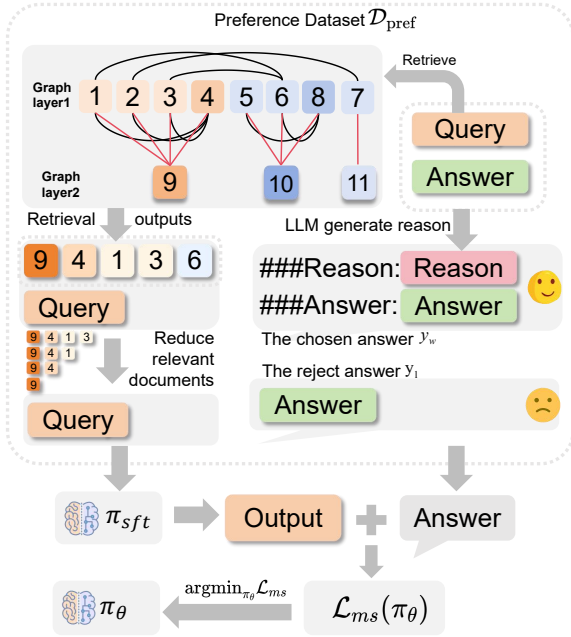


Figure 3: The human preference alignment training process with MS loss.

To this end, we first calculate the query embedding for subsequent small document matching and ranking:

$$e_q = \text{EMBED}(q) \quad (5)$$

Next, we calculate the similarity between the query embedding and all documents at each layer. The top k most similar documents are then selected:

$$\text{results} = \max_{D_{\text{layer}} \in \text{layers}} \text{sim}(e_q, D_{\text{layer}}) \quad (6)$$

where the query embedding e_q is compared with each document D_{layer} across all layers. The similarity function sim measures the semantic closeness between the query and the document at each layer. The k documents with the highest similarity scores are selected as the most relevant results. This method enables more efficient retrieval by leveraging the multi-layer structure of the documents.

Preferred Dataset Preparation. First, unlike previous works (Wang et al., 2024) use GPT (OpenAI et al., 2024) to generate train data, we yield reasoning explanations using the query and answer with multiple small-scale LLMs (i.e. Qwen2.5-7B (Qwen et al., 2025), LLaMA3-8B (Grattafiori et al., 2024) and Mistral-8B (Mistral AI, 2025) with Chain of Thought (CoT) (Wei et al., 2023). Note that our method of synthesizing data consumes very few resources compared with the previous direct use of large-scale GPT-4 to generate data in DPO.

Second, for each query q , we expect it to have different types and relevant contexts C to construct more diverse samples for learning. Therefore, we use $k \in [0, \text{len}(C)]$ value to select contexts ranging from strongly correlated to weakly correlated: $C_1 = [\text{doc}_9]$, $C_2 = [\text{doc}_9, \text{doc}_4]$, $C_3 = [\text{doc}_9, \text{doc}_4, \text{doc}_1] \dots$ as illustrated in Figure 3. We define two types of outputs: the positive item y_w , which includes both the reasoning process and final answer, and the negative item y_l , which only includes the answer in the training set. The overall data format is as follows:

$$\mathcal{D}_{\text{pref}} = \{(q^{(i)}, C^{(i)}, y_w^{(i)}, y_l^{(i)}) \mid i \in \{1, 2, \dots, N\}\} \quad (7)$$

Where N denotes the number of queries in the training set. An example of train data is in Appendix L.

Preference Alignment Training. With the above two types of data, we hope that our model is more inclined to generate answers that are reasonable and well-founded, that is, answers that are human-preferred. Building on the success of Direct Preference Optimization (DPO) (Rafailov et al., 2024) with reinforcement learning in the training of LLMs, we propose to seek the mode of real distribution with the mode-seeking loss function (MS), rather than mean-seeking in DPO (Chan et al., 2022; Ji et al., 2024). It is defined as:

$$\mathcal{L}_{\text{ms}}(\pi_\theta) = \mathbb{E}_{x \sim \mathcal{D}_{\text{pref}}} \mathbb{E}_{\pi_{\text{sft}}} (\mathbf{y}_{1:K} | x) \left[\mathbb{D}_{\text{KL}}(p_{f_\theta}(\cdot | \mathbf{y}_{1:K}, x) \| p_{r_\phi}(\cdot | \mathbf{y}_{1:K}, x)) \right] \quad (8)$$

where $\mathcal{D}_{\text{pref}}$ denotes the preference dataset which contains human-labeled preference data. $\pi_{\text{sft}}(y_{1:K} | x)$ indicates the supervised fine-tuned (SFT) policy and π_{sft} means the probability distribution of responses given a prompt x after the language model has undergone supervised fine-tuning. $p_{f_\theta}(\cdot | y_{1:K}, x)$ represents the empirical distribution based on the model policy. $p_{r_\phi}(\cdot | y_{1:K}, x)$ is the empirical distribution based on the reward model.

4 Experiments

We conduct experiments to evaluate our method on diverse QA tasks against baselines.

4.1 Experimental Setting

Datasets. We evaluate various QA datasets, roughly divided into: 1) **GenerativeQA:** QASPER (Dasigi et al., 2021), evaluated with the

	QASPER	QuALITY	RiddleSense	PubMedQA	MedQA	MedMcQA
	RAPTOR (Sarathi et al., 2024)					
LLaMa 8B	0.3657	45.62	49.66	58.40	53.10	50.84
	LightGraphRAG (Guo et al., 2024)					
LLaMa 8B	0.3585	45.82	50.83	49.00	45.18	50.91
	Reward-RAG (Nguyen et al., 2024)					
GPT-3.5-turbo	-	-	-	69.20	59.20	52.40
GPT-4-0613	-	-	-	70.80	64.50	57.40
	LLaMa 8B (Grattafiori et al., 2024)					
Basic LLM	0.1040	32.10	62.78	49.60	60.17	50.01
Basic RAG	0.3599	41.73	60.24	68.80	57.34	50.35
GraphMPA (ours)	0.3775	47.05	73.65	73.00	66.54	64.28
	Qwen 7B (Qwen et al., 2025)					
Basic LLM	0.088	41.54	65.95	28.60	52.00	53.36
Basic RAG	0.2654	47.32	65.95	50.67	50.82	55.15
GraphMPA (ours)	0.3734	47.64	71.79	71.92	61.96	57.61
	Mistral 8B (Mistral AI, 2025)					
Basic LLM	0.1135	35.32	59.39	47.40	54.28	58.07
Basic RAG	0.3228	43.64	64.29	66.80	56.64	63.93
GraphMPA (ours)	0.3873	51.76	73.92	72.82	68.66	67.06

Table 1: Performance comparison of various models and approaches across different QA datasets. The marker '-' denotes the results unavailable in public reports.

Dataset	Docs	TrainQAs	TestQAs	Avg Tokens
QASPER	6724	5000	3328	358
QuALITY	5772	4600	2128	365
PubMed	1500	5000	500	398
Riddle	3510	5000	1021	150
MedQA	3765	5000	1273	194
MedMC	16777	5000	4183	175

Table 2: The dataset statistics include the number of small documents, train QA pairs, test QA pairs, and average document tokens.

ROUGE score (Recall-Oriented Understudy for Gisting Evaluation)(Lin, 2004; Ganesan, 2018); 2) **ChoiceQA**: QuALITY (Pang et al., 2022) and RiddleSense (Lin et al., 2021a), evaluated with standard *Accuracy*; 3) **MedicalQA**: PubMedQA (Jin et al., 2019), MedQA (Jin et al., 2020) and MedMCQA (Pal et al., 2022), evaluated with *MIRAGE* (Medical Information Retrieval-Augmented Generation Evaluation) (Xiong et al., 2024) *. The statistical summary can refer to Table 2, where TrainQAs represents the original split of the training set in each dataset. We use TrainQAs to automatically synthesize 20000 samples for all datasets to conduct subsequent preference alignment training.

Baselines and Implementation Details. We compare the following representative SOTA on RAG and QA: Standard RAG (Lewis et al., 2021; Gao et al., 2024), RAPTOR (Sarathi et al.,

2024), LightGraphRAG (Guo et al., 2024) and Reward-RAG (Nguyen et al., 2024). In our approach, we use three backbones: LLaMa-3.1-8B-Instruct (Grattafiori et al., 2024; Patterson et al., 2022), (i.e., LLaMa 8B), Qwen2.5-7B-Instruct (Qwen et al., 2025) (i.e., Qwen 7B), and Mistral-8B-Instruct-2410 (Mistral AI, 2025) (i.e., Mistral 8B). We implement training using the trl library, configuring the learning rate to 1e-5, the batch size to 4, and training for 2 epochs.

4.2 Main Results

The experimental results presented in Table 1 provide a thorough comparison of our proposed GraphMPA with several representative SOTA across various datasets. From this table, we can find that:

It is necessary to design our GraphMPA. With the same backbone LLaMa 8B, two competitive methods RAPTOR and LightGraphRAG are inferior to our GraphMPA. This is mainly because although RAPTOR models external knowledge in a hierarchical tree, it ignores the associations among small documents in each layer (sibling and cousin nodes). Meanwhile, although LightGraphRAG adopts a graph structure based on extracted entities and their relationships as Edge et al. (2024), it overly relies on the ability to extract entities and relationships and may lose document information without entities. Additionally, Reward-RAG with very large-scale LLMs (GPT-3.5 and GPT-4), performs better than RAPTOR and LightRAG though,

*<https://github.com/mirage-project/mirage>

	QASPER	QuALITY	RiddleSense	PubMedQA	MedQA	MedMcQA
w/o Summarization	0.3599	41.73	60.24	68.80	57.34	50.35
	↓ 0.0176	↓ 5.32	↓ 13.41	↓ 4.2	↓ 9.2	↓ 13.93
w/o Retrieval	0.1040	32.10	62.78	49.60	60.17	50.01
	↓ 0.2735	↓ 14.95	↓ 10.87	↓ 23.4	↓ 6.37	↓ 14.27
w/o Training	0.3694	46.65	71.79	71.40	63.47	64.16
	↓ 0.0081	↓ 0.4	↓ 1.86	↓ 1.6	↓ 3.07	↓ 0.12
w/ DPO	0.3599	46.06	73.20	71.60	64.28	64.16
	↓ 0.0176	↓ 0.99	↓ 0.45	↓ 1.4	↓ 2.26	↓ 0.12
GraphMPA LLaMa 8B	0.3775	47.05	73.65	73.00	66.54	64.28

Table 3: The ablation study results examine the impact of removing the retrieval, summary, and training components from our model. Performance is evaluated across six datasets: QASPER, QuALITY, RiddleSense, PubMedQA, MedQA and MedMcQA. The results show that removing the retrieval component causes the largest drop in performance, followed by the removal of the summary and training components. The full model (ours) achieves the highest performance on all datasets, highlighting the importance of each component in enhancing model effectiveness.

it still indicates worse performance than our approach. This may be due to outdated human preference optimization strategies like RLHF. This suggests that we should design a comprehensive graph framework and employ an advanced human preference alignment mechanism.

RAG-based technique is worth further exploration as our GraphMPA. Compared to the basic LLMs with different backbones, the addition of RAG technology improves performance. This indicates that RAG is indeed effective for QA. Besides, as our GraphMPA with improved the RAG-based technique, the performance has further improved. This suggests that RAG-based technology is worth further exploration and improvement.

Our GraphMPA performs robustly in different frameworks. We evaluate GraphMPA using three different LLMs: LLaMa 8B, Qwen 7B, and Mistral 8B. Among these, the 8B LLMs significantly outperform prior models, while Qwen 7B surpasses all other models, excluding GPT-4. This improvement is largely attributed to the increased parameter size of the 8B models, which leads to better performance. These results indicate that GraphMPA exhibits strong robustness across different model architectures.

4.3 Analysis and Discussion

Ablation Study. Table 3 displays the results of removing the key components in our approach. From this table, we can see that removing any module will result in significant performance degradation. This indicates that every module designed in our method is important. Among them, the removal of the summarization and retrieval modules results in the most severe performance degradation. This

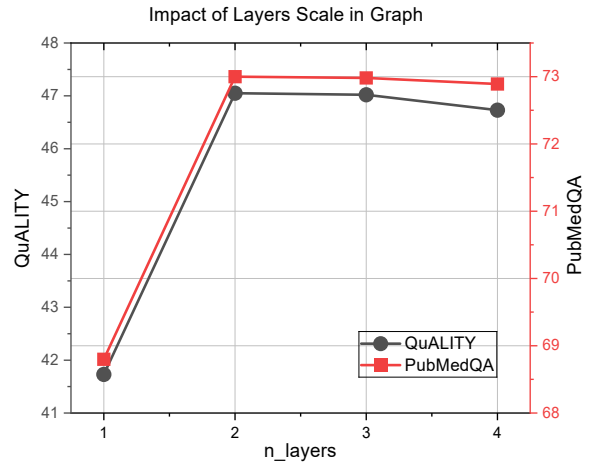


Figure 4: Performance comparison of our GraphMPA on both QuALITY and PubMedQA, as graph layers *Layers* changes.

indicates that it is crucial to effectively represent external knowledge and retrieve relevant knowledge. Therefore, this paper designs a comprehensive hierarchical graph framework to better accomplish representation and retrieval.

Impact of *Layers Scale in Graph*. Figure 4 displays the performance of our GraphMPA as graph layers increasing, where $n_layers = 1$ functions as a basic RAG. From this figure, we can observe that the performance on each dataset first improves at layer 2 and then tends to stabilize or even slightly decrease. This suggests the effectiveness of our designed hierarchical graph with community summarization. However, excessive summarization when building a deep graph does not always lead to better results since too many iterations of summarization may render the document abstract and less meaningful.

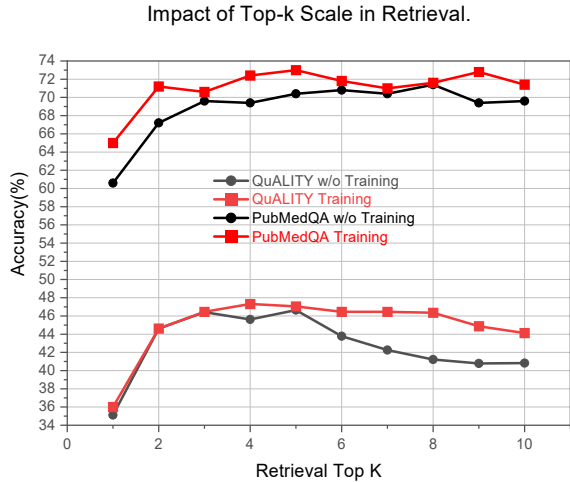


Figure 5: Performance comparison on both trained and untrained models with varying values of retrieval top- k regarding both QuALITY and PubMedQA datasets.

Impact of Top- k Scale in Retrieval. In Figure 5, as the number of retrieved top k increases, the performance of different models generally improves first and then decreases on each dataset. This is mainly because the number of unrelated contexts increases when k increases, which affects the uncertainty of the model’s generation. Additionally, we observe that the accuracy of the untrained model decreases fast, while the trained model decreases slowly. This indicates that the training allows the model to extract relevant information and discriminate both related and unrelated contexts for better answers.

Moreover, statistical analysis of the distribution of top- k documents across different graph layers is available in Appendix H. And the importance of ranking in Appendix I.

Effectiveness of mode-seeking in generation.

The log probability $\log \pi(y|x)$ is used in alignment tasks to compute the KL divergence or other optimization objectives, ensuring that the optimized policy $\pi(y|x)$ remains consistent with human preferences (Chan et al., 2022; Ji et al., 2024). In other words, the higher the value, the more in line with human preferences. Therefore, for each sample in PubMedQA, we calculate the log probabilities of $\pi_{ms}(y|x)$, $\pi_{dpo}(y|x)$ and $\pi_{sft}(y|x)$, as illustrated in Figure 6. From this figure, we can see that the median of our MS strategy outperforms the other two, suggesting a better human preference alignment. From the perspective of sample distribution, we can observe that each data point of our MS loss is concentrated in a small range to seek the main

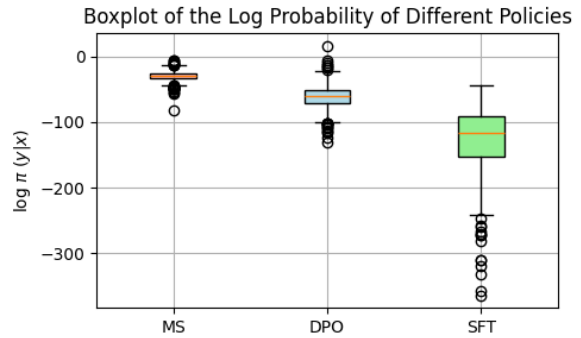


Figure 6: The log probabilities on PubMedQA with the trained model using MS, DPO and non-trained SFT model.

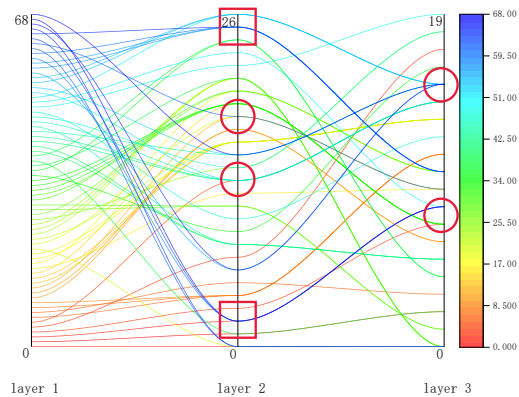


Figure 7: An example of community summarization process in our graph building from QuALITY. Layer 1 consists of 69 small documents, Layer 2 is the community summary (0-26) derived from Layer 1, and Layer 3 is the community summary (0-19) derived from Layer 2.

mode, rather than being discrete like DPO data points, striving for a comprehensive (mean) mode. Meanwhile, SFT discretizes to more remote areas and cannot even achieve the average mode.

Case Study. Figure 7 illustrates a real example from QuALITY, where each line represents the direction in which each point gathers from left to right. From this figure, we can see that we can see that on the second and third layers, whether they were previously far apart or close together, they will gather together according to the community. This indicates that our approach can effectively organize and understand external knowledge, meaning that our summarization-based graph structure is effective. This corresponding textual content, comparison of output results from different models (Sec. K.1), and comprehensive comparison of graph construction using different methods (Sec. K.2) can be found in the appendix.

5 Conclusion

This work explores a summarization-based hierarchical graph to comprehensively extract both low- and high-level information from external knowledge for answering questions (QA). Moreover, we utilize small-scale LLMs to automatically synthesize data on human preferences and employ mode-seeking loss to capture the main patterns of the optimal policy, thereby better achieving the output of human preferences.

6 Acknowledgements

This research/project is supported by the Ministry of Education, Singapore under its MOE Academic Research Fund Tier 2 (MOE-T2EP20123-0005: “Neurosymbolic AI for Commonsense-based Question Answering in Multiple Domains”), by the National Natural Science Foundation of China grant (NSFC No. 62206193 and No. 62376178), and by the General Research Fund (GRF) project sponsored by the Research Grants Council Hong Kong (Project No. 15611021).

7 Limitations

This study has several limitations, primarily due to computational resource constraints. Specifically, we were unable to utilize LLMs with more than 8 billion parameters or state-of-the-art (SOTA) models such as GPT-4. Consequently, some prior studies could not be fully reproduced. Nevertheless, our method achieves superior performance on several benchmark datasets compared to approaches leveraging GPT-4 as the backbone model, demonstrating its effectiveness despite these limitations.

References

Tahani H Alwaneen, Aqil M Azmi, Hatim A Aboalsamh, Erik Cambria, and Amir Hussain. 2022. Arabic question answering system: A survey. *Artificial Intelligence Review*, 55(1):207–253.

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2024. SELF-RAG: LEARNING TO RETRIEVE, GENERATE, AND CRITIQUE THROUGH SELF-REFLECTION.

Scott Barnett, Stefanus Kurniawan, Srikanth Thudumu, Zach Brannelly, and Mohamed Abdelrazek. 2024. Seven failure points when engineering a retrieval augmented generation system. *Preprint*, arXiv:2401.05856.

Erik Cambria, Bjoern Schuller, Bing Liu, Haixun Wang, and Catherine Havasi. 2013. Statistical approaches

to concept-level sentiment analysis. *IEEE Intelligent Systems*, 28(3):6–9.

- Alan Chan, Hugo Silva, Sungsu Lim, Tadashi Kozuno, A. Rupam Mahmood, and Martha White. 2022. Greedification operators for policy optimization: Investigating forward and reverse KL divergences. *J. Mach. Learn. Res.*, 23:253:1–253:79.
- Paul Chandler and John Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and instruction*, 8(4):293–332.
- Jianlv Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. 2024. Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. *Preprint*, arXiv:2402.03216.
- Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2023. Deep reinforcement learning from human preferences. *Preprint*, arXiv:1706.03741.
- Pradeep Dasigi, Kyle Lo, Iz Beltagy, Arman Cohan, Noah A. Smith, and Matt Gardner. 2021. A dataset of information-seeking questions and answers anchored in research papers. *Preprint*, arXiv:2105.03011.
- Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. 2024. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. *Preprint*, arXiv:2404.16130.
- Kavita Ganesan. 2018. Rouge 2.0: Updated and improved measures for evaluation of summarization tasks. *Preprint*, arXiv:1803.01937.
- Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey. *Preprint*, arXiv:2312.10997.
- Aaron Grattafiori, Abhimanyu Dubey, and Abhinav Jauhri. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.
- Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. 2024. LightRAG: Simple and Fast Retrieval-Augmented Generation. *arXiv preprint*. ArXiv:2410.05779 [cs].
- Zihao Huang, Rui Mao, Xiaobao Wu, Kai He, Xulang Zhang, and Erik Cambria. 2025. Fundamental algorithms for retrieval augmented generation: A survey. *IEEE Transactions on Knowledge and Data Engineering*.
- Haozhe Ji, Cheng Lu, Yilin Niu, Pei Ke, Hongning Wang, Jun Zhu, Jie Tang, and Minlie Huang. 2024. Towards efficient exact optimization of language model alignment. In *Proceedings of the 41st International Conference on Machine Learning, ICML’24*. JMLR.org.

- Di Jin, Eileen Pan, Nassim Oufattole, Wei-Hung Weng, Hanyi Fang, and Peter Szolovits. 2020. [What disease does this patient have? a large-scale open domain question answering dataset from medical exams](#). *Preprint*, arXiv:2009.13081.
- Qiao Jin, Bhuwan Dhingra, Zhengping Liu, William W. Cohen, and Xinghua Lu. 2019. [Pubmedqa: A dataset for biomedical research question answering](#). *Preprint*, arXiv:1909.06146.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. [Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks](#). *Preprint*, arXiv:2005.11401.
- Bill Yuchen Lin, Ziyi Wu, Yichi Yang, Dong-Ho Lee, and Xiang Ren. 2021a. [Riddlesense: Reasoning about riddle questions featuring linguistic creativity and commonsense knowledge](#). *Preprint*, arXiv:2101.00376.
- Chin-Yew Lin. 2004. [ROUGE: A package for automatic evaluation of summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Chu-Cheng Lin, Aaron Jaech, Xin Li, Matthew R. Gormley, and Jason Eisner. 2021b. [Limitations of autoregressive models and their alternatives](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 5147–5173. Association for Computational Linguistics.
- Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2023. [Lost in the middle: How language models use long contexts](#). *Preprint*, arXiv:2307.03172.
- Qian Liu, Xiubo Geng, Yu Wang, Erik Cambria, and Daxin Jiang. 2024. [Disentangled retrieval and reasoning for implicit question answering](#). *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7804–7815.
- Mistral AI. 2025. [Un Ministral, des Ministraux](#). <https://mistral.ai/news/ministraux/>. Accessed: 2025-01-19.
- Thang Nguyen, Peter Chin, and Yu-Wing Tai. 2024. [Reward-RAG: Enhancing RAG with Reward Driven Supervision](#). *Preprint*, arXiv:2410.03780.
- Luca Oneto, Federica Bisio, Erik Cambria, and Davide Anguita. 2016. [Statistical learning theory and ELM for big social data analysis](#). *IEEE Computational Intelligence Magazine*, 11(3):45–55.
- OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, and Lama Ahmad. 2024. [Gpt-4 technical report](#). *Preprint*, arXiv:2303.08774.
- Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. [Training language models to follow instructions with human feedback](#). *Preprint*, arXiv:2203.02155.
- Ankit Pal, Logesh Kumar Umapathi, and Malaikannan Sankarasubbu. 2022. [Medmcqa : A large-scale multi-subject multi-choice dataset for medical domain question answering](#). *Preprint*, arXiv:2203.14371.
- Richard Yuanzhe Pang, Alicia Parrish, Nitish Joshi, Nikita Nangia, Jason Phang, Angelica Chen, Vishakh Padmakumar, Johnny Ma, Jana Thompson, He He, and Samuel R. Bowman. 2022. [Quality: Question answering with long input texts, yes!](#) *Preprint*, arXiv:2112.08608.
- David Patterson, Joseph Gonzalez, Urs Hölzle, Quoc Le, Chen Liang, Lluís-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2022. [The carbon footprint of machine learning training will plateau, then shrink](#). *Preprint*, arXiv:2204.05149.
- Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. 2025. [Qwen2.5 technical report](#). *Preprint*, arXiv:2412.15115.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. [Direct preference optimization: Your language model is secretly a reward model](#). *Preprint*, arXiv:2305.18290.
- Parth Sarthi, Salman Abdullah, Aditi Tuli, Shubh Khanna, Anna Goldie, and Christopher D. Manning. 2024. [RAPTOR: Recursive Abstractive Processing for Tree-Organized Retrieval](#). *Preprint*, arXiv:2401.18059.
- John Sweller. 1988. [Cognitive load during problem solving: Effects on learning](#). *Cognitive Science*, 12(2):257–285.
- Jiaming Tang, Yilong Zhao, Kan Zhu, Guangxuan Xiao, Baris Kasikci, and Song Han. 2024. [Quest: Query-aware sparsity for efficient long-context llm inference](#). *Preprint*, arXiv:2406.10774.
- Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. 2019. [From louvain to leiden: guaranteeing well-connected communities](#). *Scientific reports*, 9(1):1–12.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. [Improving text embeddings with large language models](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11897–11916, Bangkok, Thailand. Association for Computational Linguistics.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. 2023. [Chain-of-thought prompting elicits reasoning in large language models](#). *Preprint*, arXiv:2201.11903.

Jincenzi Wu, Zhuang Chen, Jiawen Deng, Sahand Sabour, Helen Meng, and Minlie Huang. 2024. [COKE: A cognitive knowledge graph for machine theory of mind](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 15984–16007. Association for Computational Linguistics.

Guangzhi Xiong, Qiao Jin, Zhiyong Lu, and Aidong Zhang. 2024. [Benchmarking retrieval-augmented generation for medicine](#). *Preprint*, arXiv:2402.13178.

Shicheng Xu, Liang Pang, Mo Yu, Fandong Meng, Huawei Shen, Xueqi Cheng, and Jie Zhou. 2024. [Unsupervised Information Refinement Training of Large Language Models for Retrieval-Augmented Generation](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 133–145. Association for Computational Linguistics.

Penghao Zhao, Hailin Zhang, Qinhan Yu, Zhengren Wang, Yunteng Geng, Fangcheng Fu, Ling Yang, Wentao Zhang, Jie Jiang, and Bin Cui. 2024. [Retrieval-augmented generation for ai-generated content: A survey](#). *Preprint*, arXiv:2402.19473.

A A Simple Example of Graph Construction

Figure 8 illustrates an example of graph construction provided using two documents about cities in China, sourced from Wikipedia. First, we summarize each document. Next, we split the documents into smaller segments. Then, we embed these segments into dense vectors and calculate the similarity between them. We treat the segments as nodes, and based on their similarity, we establish edges between the nodes. Using this graph, we apply community detection algorithms to identify communities, which are then summarized to extract high-level nodes.

B Details of Mode-seeking Preference Alignment

We provide a simple justification for why our mode-seeking loss with small-scale auto-synthesized training data is more satisfying than the mean-seeking loss of traditional DPO with large-scale LLM-generated training data.

Optimal Policy π^* is defined as follows:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{x \sim D, y \sim \pi(y|x)} [r_{\phi}(x, y)] - \beta \cdot D_{KL}(\pi(y|x) || \pi_{\text{sft}}(y|x)) \quad (9)$$

where $r_{\phi}(x, y)$ is the reward model, $\pi_{\text{sft}}(y|x)$ is the initial policy, and β is a regularization parameter.

MS minimizes the reverse KL divergence:

$$\pi_{ms} = \arg \max_{\pi} D_{KL}(\pi(y|x) || \pi^*(y|x)) \quad (10)$$

DPO minimizes the forward KL divergence:

$$\pi_{dpo} = \arg \max_{\pi} D_{KL}(\pi^*(y|x) || \pi(y|x)) \quad (11)$$

The reverse KL divergence is defined as follows:

$$\begin{aligned} & D_{KL}(\pi(y|x) || \pi^*(y|x)) \\ &= \mathbb{E}_{y \sim \pi(y|x)} \left[\log \frac{\pi(y|x)}{\pi^*(y|x)} \right] \end{aligned} \quad (12)$$

Mode-Seeking: Minimizing reverse KL divergence $D_{KL}(\pi(y|x) || \pi^*(y|x))$ encourages $\pi(y|x)$ to concentrate on the main modes of $\pi^*(y|x)$.

The forward KL divergence is defined as follows:

$$\begin{aligned} & D_{KL}(\pi^*(y|x) || \pi(y|x)) \\ &= \mathbb{E}_{y \sim \pi(y|x)} \left[\log \frac{\pi^*(y|x)}{\pi(y|x)} \right] \end{aligned} \quad (13)$$

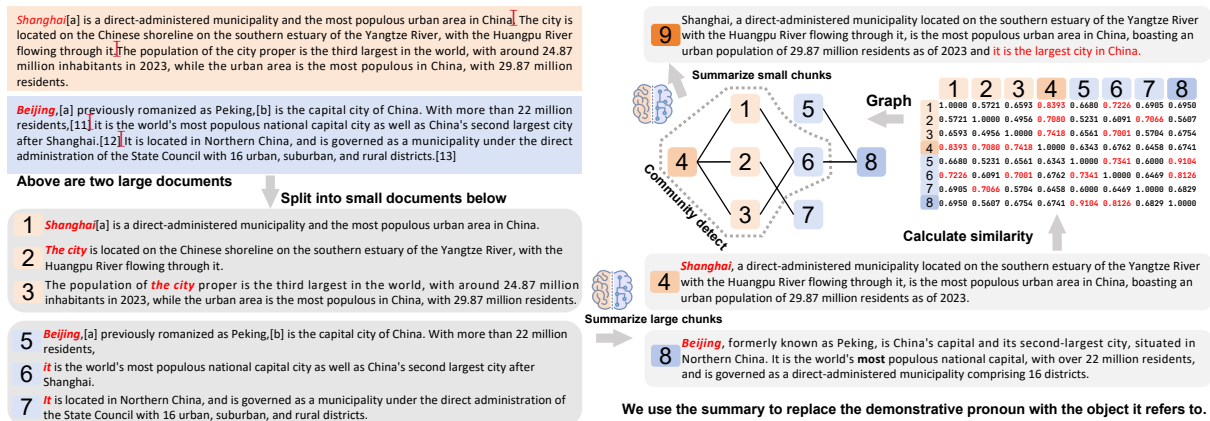


Figure 8: The figure shows two descriptions of the cities Shanghai and Beijing from wikipedia. As part of the basic RAG process, we split the large documents into small chunks (documents) 1 ~3 5 ~7, embed them and store them in a vector database. During retrieval, the basic RAG process may return incorrect references (Query "Shanghai" get 6). To address this, we apply abstractive summarization to large documents to eliminate misrepresentation. Furthermore, we build a graph \mathcal{G} based on the similarity between documents. Using this graph, we group similar documents into the same community and apply abstractive summarization to obtain broader insights.

Mean-Seeking: Minimizing forward KL divergence $D_{KL}(\pi^*(y|x) || \pi(y|x))$ encourages $\pi^*(y|x)$ to cover the entire support of $\pi^*(y|x)$, potentially leading to a distribution that is less concentrated on the main modes (Ji et al., 2024).

Minimizing the reverse KL divergence $D_{KL}(\pi(y|x) || \pi^*(y|x))$ causes π_{ms} to concentrate on the main modes of π^* because it penalizes π_{ms} for assigning high probability to regions where π^* has low probability. This mode-seeking behavior ensures that π_{ms} closely approximates the main modes of π^* , making it an effective way to optimize policies in alignment tasks.

Simply put, building on the success of Direct Preference Optimization (DPO) (Rafailov et al., 2024) based on reinforcement learning in the training of LLMs, we propose to seek the mode distribution of the optimal policy (Ji et al., 2024) with the mode-seeking loss function (MS) as follows:

$$\mathcal{L}_{ms}(\pi_\theta) = \mathbb{E}_{x \sim \mathcal{D}_{\text{pref}}} \mathbb{E}_{\pi_{\text{sft}}} (\mathbf{y}_{1:K} | x) \left[\mathbb{D}_{\text{KL}}(p_{f_\theta}(\cdot | \mathbf{y}_{1:K}, x) || p_{r_\phi}(\cdot | \mathbf{y}_{1:K}, x)) \right] \quad (14)$$

where $\mathcal{D}_{\text{pref}}$ is the preference dataset which contains human-labeled preference data. Each entry includes a prompt x and a set of responses $(y_{1:k})$, where y_1, \dots, y_k are sampled responses and Labels indicating which responses are preferred or rejected.

$\pi_{\text{sft}}(y_{1:K} | x)$ is the supervised fine-tuned (SFT) policy π_{sft} is the probability distribution of re-

sponses given a prompt x after the language model has undergone supervised fine-tuning.

$p_{f_\theta}(\cdot | y_{1:K}, x)$ is the empirical distribution based on the model policy. This distribution is computed over the sampled responses $y_{1:K}$ and reflects the model's current policy π_θ . Represents the model's belief over the sampled responses $y_{1:K}$.

$p_{r_\phi}(\cdot | y_{1:K}, x)$ is the empirical distribution based on the reward model. This distribution reflects the reward model's scoring over the sampled responses $y_{1:K}$, based on human preferences. Represents the "ideal" distribution of responses based on human preferences.

$\mathbb{D}_{\text{KL}}(p_{f_\theta} || p_{r_\phi})$: Kullback-Leibler (KL) divergence, which aligns the model's output distribution p_{f_θ} with the distribution defined by the reward model p_{r_ϕ} .

$\mathbb{E}_{x \sim \mathcal{D}_{\text{pref}}}$ ensures generalization of the policy π_θ to the entire dataset.

$\mathbb{E}_{\pi_{\text{sft}}(y_{1:K} | x)}$ captures the effect of sampling different response sets on the loss.

The MS loss aligns the model's policy π_θ with human preferences by minimizing the KL divergence between the model's empirical distribution (p_{f_θ}) and the reward model's empirical distribution (p_{r_ϕ}) over sampled responses, averaged across all prompts in the dataset. It effectively optimizes the policy while maintaining computational efficiency.

C Details of Baselines

C.1 Baselines

- Standard RAG (Lewis et al., 2021; Gao et al., 2024) is the most classic method for retrieving external knowledge to augment the performance of LLMs.
- RAPTOR (Sarathi et al., 2024) builds a summary tree with text chunk embedding and clustering and retrieves from the tree to offer LLM additional information from the outer.
- LightRAG (Guo et al., 2024) incorporates graph structures into text indexing and retrieval processes and allows the system to remain effective and responsive in rapidly changing data environments compared to the Graph RAG (Edge et al., 2024) which builds a graph using LLM to extract entity-relation triples from text and makes summaries from that to build a higher-level graph, then retrieves from these graphs to provide LLM outer knowledge.
- Reward-RAG (Nguyen et al., 2024) aligns RAG with human preferences by integrating a reward model to train a higher-performing embedding model.

D Details of Datasets

D.1 Datasets

QASPER (Dasigi et al., 2021) consists of 5,049 questions over 1,585 Natural Language Processing papers that are often read to seek information present in the full text to answer specific questions.

QuALITY (Pang et al., 2022), Question Answering with Long Input Texts, Yes! is a dataset to enable training and testing models on long-document comprehension, which consists of multiple-choice QA with context passages in English.

RiddleSense (Lin et al., 2021a) consists of 5.7k examples and aims to build and test the model’s reasoning about riddle questions consisting of questions and multiple choice.

PubMedQA (Jin et al., 2019) is a biomedical question-answering (QA) dataset collected from PubMed abstracts that have 1k expert annotations. The task of PubMedQA is to answer research questions with yes/no/maybe.

MedQA (Jin et al., 2020) collected from the professional medical board exams, which contain

12,723 questions in English, and each question is accompanied by its answer and several options with an alpha index.

MedMCQA (Pal et al., 2022) is a large-scale, Multiple-Choice Question Answering (MCQA) dataset built from 194k high-quality medical exams, designed to address real-world medical entrance exam questions. Each sample contains a question, correct answer(s), and other options.

E Details of Evaluation Metrics

E.1 Evaluation Metrics

We evaluate text generation outputs using the ROUGE score (Recall-Oriented Understudy for Gisting Evaluation)(Lin, 2004; Ganesan, 2018) F1, a standard metric for assessing the quality of generated text by comparing n-gram overlaps between the predicted and reference outputs. For tasks involving discrete responses, such as yes/no, A/B/C/D, or 1/2/3/4 choices, we report accuracy as the evaluation metric.

Additionally, for evaluating our model’s performance on medical question-answering datasets, we employ MIRAGE (Medical Information Retrieval-Augmented Generation Evaluation) (Xiong et al., 2024) *, a specialized metric designed for assessing retrieval-augmented generation models in the medical domain. We apply MIRAGE to evaluate our performance on three benchmark datasets: PubMedQA, MedQA, and MeMCDQA.

F Detailed Metrics of QASPER

Overall, our GraphMPA consistently outperforms all baselines in terms of F1, Precision, Recall, BLEU, and Meteor scores, demonstrating the effectiveness of our approach.

G Impact of τ Value in Graph.

As the value of τ increases, the number of edges decreases. As illustrated in Figure 9, performance reaches its maximum around a threshold of 0.5. Both metrics exhibit an upward trend as the threshold increases from 0.0 to approximately 0.5, peaking at this threshold before declining as the threshold exceeds 0.5. These observations indicate that both QuALITY and PubMedQA attain optimal values at a threshold of around 0.5. Performance for both metrics declines when the threshold is either too low or too high. The fluctuation induced by τ is

*<https://github.com/mirage-project/mirage>

QASPER	F1	Precision	Recall	BLEU	Meteor
RAPTOR (Sarathi et al., 2024)					
LLaMa 8B	0.3657	0.4660	0.3009	0.1158	0.3338
LightGraphRAG (Guo et al., 2024)					
LLaMa 8B	0.3585	0.5781	0.2598	0.1262	0.3834
LLaMa 8B (Grattafiori et al., 2024)					
Basic LLM	0.1040	0.1612	0.0767	0.0221	0.0996
Basic RAG	0.3599	0.5307	0.2800	0.1723	0.3519
GraphMPA (ours)	0.3775	0.5645	0.2835	0.1745	0.3982
Qwen 7B (Qwen et al., 2025)					
Basic LLM	0.0881	0.1436	0.0813	0.0174	0.0913
Basic RAG	0.2654	0.6645	0.2369	0.1675	0.3321
GraphMPA (ours)	0.3734	0.6420	0.2632	0.1885	0.3921
Mistral 8B (Mistral AI, 2025)					
Basic LLM	0.1135	0.1784	0.0832	0.0238	0.1074
Basic RAG	0.3228	0.5919	0.2219	0.1294	0.3824
GraphMPA (ours)	0.3873	0.6314	0.2793	0.1647	0.4135

Table 4: The performance of various models on QASPER is evaluated across several metrics, including F1, Precision, Recall, BLEU, and Meteor. We present results for the RAPTOR (Sarathi et al., 2024), LightGraphRAG (Guo et al., 2024), LLaMa 8B (Grattafiori et al., 2024), Qwen 7B (Qwen et al., 2025), and Mistral 8B (Mistral AI, 2025) models, with a focus on different configurations: Basic LLM, Basic RAG, and GraphMPA (our approach).

approximately 1 point, which does not significantly impact the robustness of our method.

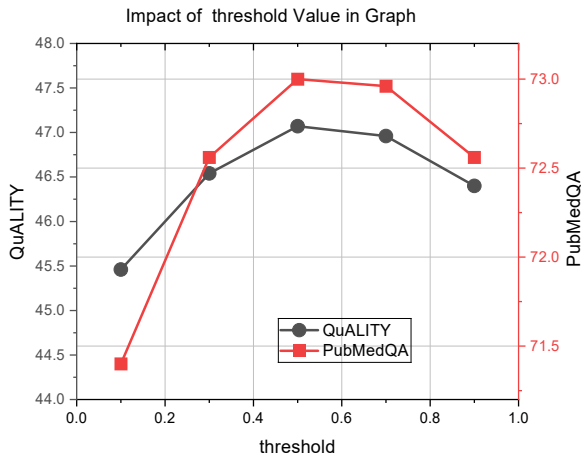


Figure 9: Impact of τ Value in Graph.

H Contributions Analysis of Layers

Statistical analysis of the top-k document distribution across different graph layers highlights the respective contributions of both low-level and high-level documents to the final output generation.

We retrieve 10 small documents (Top-k = 10) on a graph architecture with only 2 layers (n-layers = 2) for the test sets (LLM = llama3-8b).

Tables 5 and 6 show the proportion of test set samples with 10 retrieved relevant documents distributed across two layers of the graph network. For

PubMedQA	0-2	3-5	6-8	9-10
layer 1 (low-level)	0	0.032	0.73	0.238
layer 2 (high-level)	0.592	0.394	0.014	0.0

Table 5: Contributions Analysis of the distribution of top-k documents across different graph layers on PubMedQA.

QuALITY	0-2	3-5	6-8	9-10
layer 1 (low-level)	0.013	0.170	0.633	0.184
layer 2 (high-level)	0.415	0.505	0.079	0.001

Table 6: Contributions Analysis of the distribution of top-k documents across different graph layers on QuALITY.

example, 0.592 means that out of the 10 retrieved documents, 0-2 of them fall on the second layer, accounting for 59.2% of the test set samples. From 5 and 6, we can observe that the retrieved relevant documents may appear at any level. Some samples obtain more documents on the second layer than on the first layer after ranking, while others do the opposite. This indicates both low- and high-level information is important, which aligns with our motivation. Besides, the retrieved documents appear more in the first layer in general. This is mainly because the nodes in the second layer are usually fewer than those in the first layer.

I The Importance of Ranking

	PubMedQA	QuALITY
only llm w/o graph	49.60	32.10
only layer 1 (low-level)	68.80	41.73
only layer 2 (high-level)	62.60	40.23
both layers	73.00	47.05

Table 7: Generating outputs using different levels of layers.

From table 7, we can see that compared to not using a graph structure (only llm w/o graph), using only one layer of nodes (only layer 1 or layer 2) produces better results. Additionally, using only layer 1 (low-level) produces relatively better performance than using only layer 2 (high-level). Of course, using both low- and high-level information simultaneously (both layers) will achieve significant performance improvements. This indicates that: 1) It is necessary to design comprehensive graph networks and retrieval mechanisms to model external information, which is consistent with the motivation of this paper; 2) The amount of low-

level information is greater, which is consistent with the distribution we calculated above. 3) Both low-level and high-level information have their own advantages, and utilizing them can achieve optimal performance. This motivates the design of our approach.

J Details of Building Graph Algorithm

In the practical implementation, to save computational costs, we also select the top k edges by similarity measurement of the current node that connects to other nodes. The detailed process can refer to Algorithm 2.

Algorithm 2 Build Graph Algorithm

```

1: function BUILD GRAPH(text ▷ Document,
   large, small,           ▷ Output length
   k,                       ▷ Top K to build graph
   n_layers,                ▷ Depth  $L$ 
    $\tau$                        ▷ Threshold )
2:  $D_{large} \leftarrow \text{TextSplit}(text, large)$ 
3:  $S_D \leftarrow \text{LLM.summay}(D_{large}, small)$ 
4:  $S'_D \leftarrow \text{TextSplit}(D_{large}, small)$ 
5:  $D \leftarrow S_D \cup S'_D$ 
6:  $layers \leftarrow []$ 
7: while  $n\_layers > 0$  do
8:    $e_D \leftarrow \text{EMBED}(D)$ 
9:    $M \leftarrow \text{sim}(e_D, e_D^T)$ 
10:   $sims \leftarrow \max_k(M, axis \leftarrow 0)$ 
11:   $\mathcal{V} \leftarrow \arg \max_k(M, axis \leftarrow 0)$ 
12:   $\mathcal{E} \leftarrow []$ 
13:  for  $u \leftarrow 0$  to  $\text{len}(V)$  do
14:    for  $v$  in  $V[u]$  do
15:       $w \leftarrow sims[u][v]$ 
16:      if  $w \geq \tau$  then
17:         $E.append((u, v, w))$ 
18:      end if
19:    end for
20:  end for
21:   $\mathcal{G} \leftarrow (\mathcal{V}, \mathcal{E})$ 
22:   $layers.append(\mathcal{G})$ 
23:   $n\_layers \leftarrow n\_layers - 1$ 
   ▷ Prepare for the next layer
24:  if  $n\_layers > 0$  then
25:     $C \leftarrow \text{CommDetect}(\mathcal{G})$ 
26:     $D \leftarrow \text{LLM.summay}(C, small)$ 
27:  end if
28: end while
29: return layers
30: end function

```

K Case Study

K.1 A case of PubMedQA question answering

Question:

Treadmill testing of children who have **spina bifida** and are ambulatory: does peak oxygen uptake reflect maximum oxygen uptake?

Our GraphMPA:

Context:['The aim of this study was to determine whether the $Vo(2)_{peak}$ measured during an incremental treadmill test is a true reflection of the maximum oxygen uptake ($(.)Vo(2)_{max}$) in children who have **spina bifida** and are ambulatory.',

'Earlier studies have demonstrated low peak oxygen uptake in children with **spina bifida**. Low peak heart rate and low peak respiratory exchange ratio in these studies raised questions regarding the true maximal character of peak values obtained with treadmill testing...',

'Twenty children who had **spina bifida** and were ambulatory participated. ...', 'The results of this study suggest that the peak measured during an incremental treadmill test is a **valid indicator** of max in children who have **spina bifida** and are ambulatory.',

"The study was conducted in a pediatric emergency department (PED) in an urban hospital. ..",]

Reason: The study aimed to determine if peak oxygen uptake during treadmill testing in children with **spina bifida** accurately reflects their maximum oxygen uptake, considering previous studies raised questions about the true maximal character of peak oxygen uptake values obtained with treadmill testing in these children.

Answer: Yes ✓

RAPTOR:

Context: Here's a summary of the study: ****Study Background and Objective:**** The study aimed to determine whether the peak oxygen uptake ($VO2_{peak}$) measured during an incremental treadmill test is a true reflection of the maximum oxygen uptake ($VO2_{max}$) in children with **spina bifida** who are ambulatory. ****Methodology:**** A cross-sectional study was conducted with 20 children with **spina bifida** who were ambulatory. The $VO2_{peak}$ was measured during a graded treadmill exercise test...

Answer: maybe ✗

LightRAG:

Answer: Maybe ✗

The inverse equity hypothesis suggests that new health interventions are initially primarily accessed by the rich, but that inequalities narrow with diffusion to the poor. However, the applicability of this hypothesis to coverage of cancer screening in middle-income countries is uncertain.

According to the data, a study examined the relationship between overall coverage and economic inequalities in coverage of cancer screening in four middle-income countries. The results showed that economic inequalities in screening participation were substantially lower or non-existent in countries with higher aggregate coverage, for both breast cancer screening and cervical cancer screening. ...

Above are case study comparisons of answering models. The responses from three different models: Ours, RAPTOR, and LightRAG to the question from PubmedQA (Jin et al., 2019). The keywords are highlighted. Our model (highlighted in green) correctly answers "Yes", In contrast, both RAPTOR and LightRAG provide uncertain and incorrect responses ("maybe"), demonstrating the difference in performance and confidence between the models.

K.2 A case of graph constructing

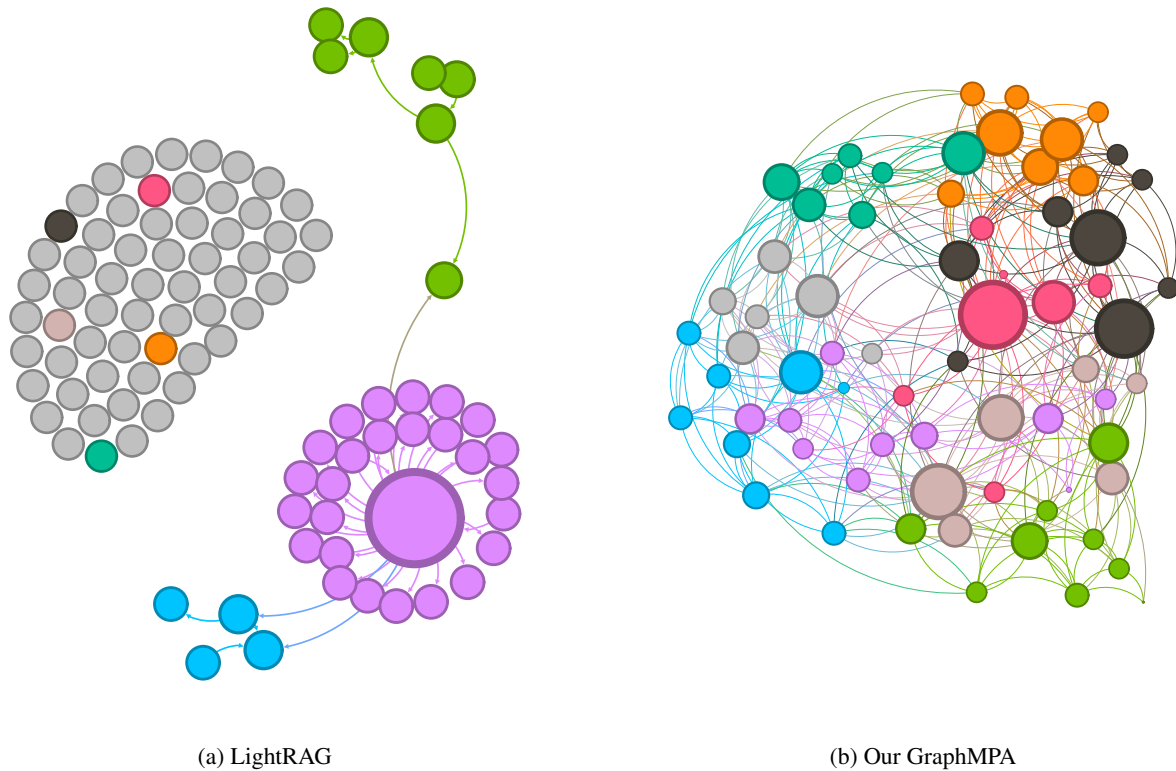


Figure 10: An example of QuALITY, LightRAG extracts nodes and edges from documents compared to our GraphMPA build graph by node similarity. GraphMPA generates nodes with rich edges while LightRAG extracts many isolated nodes.

We randomly select a paper from QuALITY to construct a graph using LightRAG and our GraphMPA. Notably, LightRAG extracts many isolated nodes, while GraphMPA generates nodes with rich edges. Then we use two methods to choose the option corresponding to the question (below) based on the graph. Due to the limited ability of LightRAG to accurately capture node relations, it struggles with relation leakage to generate the correct answer. In contrast, our GraphMPA effectively captures and generates with greater accuracy. The key reason is highlighted.

Question:

Why is Si retirement so significant to the Space Exploration Team?

Options:

1. There aren't enough working people in the world. They won't be able to find a replacement.
2. As one of two remaining spacemen, it would likely mean the defunding and shut down of the Space Exploration Team.
3. Training new spacemen is costly and time consuming. They won't have anyone else ready after him.
4. His retirement may inspire others to stop working as well, which would be hugely detrimental as most people don't feel the drive to work as is.

Process:

high_level_keywords: ["Significance of retirement", "Space Exploration Team", "Workforce impact", "Team dynamics"]

low_level_keywords: ["Si retirement", "Spacemen", "Defunding", "Shut down", "Training costs", "Replacement", "Workforce shortage", "Motivation"]

Local query uses 60 entites, 38 relations, 3 text units.

Global query uses 38 entites, 38 relations, 2 text units.

Output:

****The Significance of Si's Retirement to the Space Exploration Team****'s retirement is a significant event for the Space Exploration Team, and it's not just because he's a skilled space pilot. The team is facing a critical shortage of working people, and Si's retirement would exacerbate this issue. Here are some reasons why his retirement is so crucial to the team: 1. ****Lack of Replacement****team is struggling to find a replacement for Si, and his retirement would only worsen the situation. With a dwindling pool of skilled workers, the team would be left without a capable pilot to take on the challenges of space exploration. 2. ****Defunding and Shutdown****one of the last two remaining spacemen, Si's retirement would likely lead to the defunding and shutdown of the Space Exploration Team. The team's existence relies on the availability of skilled pilots like Si, and without him, the program would become unsustainable. 3. ****Costly and Time-Consuming Training****new spacemen is a costly and time-consuming process. The team has invested significant resources in Si's training, and his retirement would mean that those resources would be wasted. Moreover, it would take years to train a new pilot to Si's level of expertise, and even then, there's no guarantee that they would be as skilled or dedicated as Si. 4. ****Inspiring Others to Stop Working****'s retirement may also have a ripple effect, inspiring others to stop working as well. In a society where most people don't feel the drive to work, Si's decision to retire could be seen as a validation of their own apathy. This could lead to a further decline in productivity and a worsening of the already dire situation.Space Exploration Team is facing a perfect storm of challenges, and Si's retirement would be the final blow. His decision to retire is not just a personal choice; it has far-reaching consequences for the team and the future of space exploration.****The Consequences of Si's Retirement****'s retirement would have significant consequences for the Space Exploration Team, including: ****Loss of expertise****: Si's retirement would mean the loss of his expertise and experience, which would be difficult to replace. ****Defunding and shutdown****: The team's existence would be threatened, and it may be forced to shut down due to the lack of a skilled pilot. ****Inspiring others to stop working****: Si's retirement could inspire others to stop working, exacerbating the already dire situation. ****Impact on the economy****: The loss of a skilled pilot like Si would have a ripple effect on the economy, leading to a decline in productivity and a worsening of the already dire situation.Space Exploration Team is facing a critical moment, and Si's retirement is a significant factor in this crisis. The team needs to find a way to convince Si to stay on or find a suitable replacement to avoid the consequences of his retirement.",

Extract Answer: 2 ×

Global Truth: 3

Context:

["Si Pond was a space pilot, and he was retiring. He was at a bar with a woman who was clearly still in shock from meeting him. She had cried when he was given an antique watch, and was now avoiding his face. She mentioned a speech by Doctor Girard-Perregaux, who had praised Si for his accomplishments as a space pilot. Si downplayed his achievements, saying that two of his runs were only to the Moon. The woman was not amused by Si's attitude, and he offered to buy her another drink. The woman was clearly still in awe of Si, and was struggling to come to terms with the fact that he was retiring. Si, on the other hand, seemed to be taking his retirement in stride, and was more concerned with the politics of the Space Exploration department than with his own accomplishments. He mentioned that the department was in danger of being dropped by the Appropriations Committee, and that his retirement was part of a larger scheme to pressure him into taking on more trips. The woman was not impressed by Si's cynicism, and the conversation ended with him offering to buy her another drink. Key details: * Si Pond is a space pilot who is retiring.", ' There you stood, so fine and straight in your space-pilot uniform, the veteran of six exploration runs to the planets " "Well," Si said modestly, "two of my runs were only to the Moon " " and he said all those things about man's conquest of space And the dream of the stars which man has held so long And then the fact that you were the last of the space pilots The last man in the whole', " They also had a banquet for him, complete with speeches by such bigwigs of the Department of Space Exploration as Academician Lofting Gubelin and Doctor Hans Girard-Perregaux There was also somebody from the government who spoke, but he was one of those who were pseudo-elected and didn't know much about the field of space travel nor the significance of Seymour Pond's retirement Si didn't bother to remember his name", ' Si said, "Look, how about another drink " Natalie Paskov said, "Oh, I'd love to have a drink with you, Mr " "Si," Si said He motioned to the bartender with a circular twist of the hand indicating their need for two more of the same "How come you know so much about it You don't meet many people who are interested in space any more', " They hadn't figured he had enough shares of Basic to see him through decently Well, possibly he didn't, given their standards But Space Pilot Seymour Pond didn't have their standards He'd had plenty of time to think it over It was better to retire on a limited crediting, on a confoundedly limited crediting, than to take the two or three more trips in hopes of attaining a higher standard"]

Reason: Si's retirement marks the end of an era in space exploration, and he is the last of a breed of space pilots.

Answer: The significance of Si's retirement is that it marks the end of an era in space exploration and the last of a breed of space pilots.

Extract Answer: 3 ✓

Global Truth: 3

L Details of Train Data Example

Question q

Does histologic chorioamnionitis correspond to clinical chorioamnionitis?

Context C

['To evaluate the degree to which histologic chorioamnionitis, a frequent finding in placentas submitted for histopathologic evaluation, correlates with clinical indicators of infection in the mother.', 'A retrospective review was performed on 52 cases with a histologic diagnosis of acute chorioamnionitis from 2,051 deliveries at University Hospital, Newark, from January 2003 to July 2003. Third-trimester placentas without histologic chorioamnionitis (n = 52) served as controls. Cases and controls were selected sequentially. Maternal medical records were reviewed for indicators of maternal infection.', 'Histologic chorioamnionitis was significantly associated with the usage of antibiotics (p = 0.0095) and a higher mean white blood cell count (p = 0.018). The presence of 1 or more clinical indicators was significantly associated with the presence of histologic chorioamnionitis (p = 0.019).']

Chosen answer y_w

###Reason: Histologic chorioamnionitis is a reliable indicator of infection whether or not it is clinically apparent.

###Answer: yes

Rejected answer y_l

yes