Bridging Relevance and Reasoning: Rationale Distillation in Retrieval-Augmented Generation

Pengyue Jia[1]*, Derong Xu[1,2]*, Xiaopeng Li[1]*, Zhaocheng Du[3], Xiangyang Li[3]
Yichao Wang[3], Yuhao Wang[1], Qidong Liu[1], Maolin Wang[1], Huifeng Guo[3]
Ruiming Tang[3]†, Xiangyu Zhao[1]†
[1]City University of Hong Kong,
[2]University of Science and Technology of China, [3]Huawei Noah's Ark Lab
Correspondence: xianzhao@cityu.edu.hk

Abstract

The reranker and generator are two critical components in the Retrieval-Augmented Generation (i.e., RAG) pipeline, responsible for ranking relevant documents and generating responses. However, due to differences in pretraining data and objectives, there is an inevitable misalignment between the documents ranked as relevant by the reranker and those required by the generator to support query-specific answers. To bridge this gap, we propose **RADIO**, a novel and practical preference alignment framework with **RA**tionale **DI**stillati**O**n. Specifically, we first propose a rationale extraction method that leverages the reasoning capabilities of Large Language Models (LLMs) to extract the rationales necessary for answering a query. Subsequently, a rationale-based alignment process is designed to rerank documents based on the extracted rationales and fine-tune the reranker to better align the preferences. Extensive experiments conducted on three tasks across four datasets demonstrate the effectiveness and transferability of our approach. Our code is released online[1].

1 Introduction

Large Language Models (LLMs), pretrained on massive datasets, have demonstrated exceptional reasoning and text generation capabilities, as evidenced by prior research (Zhao et al., 2023; Ni et al., 2025; Wang et al., 2023b; Xu et al., 2024; Fu et al., 2023). These models also adhere to the scaling laws, exhibiting improvements in performance and intelligence as the number of parameters increases (Kaplan et al., 2020). Retrieval-augmented generation (RAG) (Xu et al., 2025a,b) builds upon these capabilities by integrating information retrieval mechanisms (Zhao et al., 2018a,b; Liu et al., 2024) with generative models, such as LLMs. This approach not only mitigates the problem of hallucination in text generation but also enhances the system's adaptability to dynamically evolving information needs, making it a robust solution for tasks requiring both accuracy and contextual relevance (Gao et al., 2023).

However, RAG pipelines typically assemble components (e.g., the reranker and generator (Fan et al., 2024)) that have been pretrained separately. Due to differences in their pretraining data and optimization objectives, these components often exhibit varying preferences, which can impact the overall effectiveness of the system. Specifically, pretrained rerankers (Xiao et al., 2023) are designed to evaluate the relevance between queries and documents. However, the documents identified as relevant under this criterion may not provide the necessary support for reasoning to derive an accurate answer to the query. Bridging this gap between the reranker's relevance measurement and the generator's reasoning requirements presents a significant challenge that must be addressed to improve the RAG pipeline's performance.

Recent studies try to address this gap by training a bridge model (Ke et al., 2024), using LLM-based scores (Zhang et al., 2024) or combining both LLM-based and retrieval-based scores (Dong et al., 2024) to fine-tune RAG components. Additionally, while some other methods are not explicitly designed for this problem, they can indirectly contribute to bridging the gap. These approaches can use response quality (Ma et al., 2023) or perplexity distillation (Izacard et al., 2023; Shi et al., 2023) as signals to fine-tune the reranker. Despite showing promise, these approaches face critical limitations: their alignment signals rely solely on the surface-level connection between the query/answer and document, failing to capture the deeper reasoning processes or more complex relationships involved.

To address the above limitation, we propose RA-

*Equal contribution

†Corresponding author

[1] https://github.com/Applied-Machine-Learning-Lab/RADIO

Findings of the Association for Computational Linguistics: ACL 2025, pages 4242–4256
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

DIO, a novel and practical preference alignment framework with rationale distillation in RAG. RADIO leverages rationale as a signal to bridge the reranker's relevance measurement with the generator's reasoning requirements for response generation. First, to efficiently extract the rationales needed to answer a query, we use the query and its ground truth answer as context and generate the rationales with LLMs. Second, to mitigate the preference misalignment between the reranker and generator while ensuring the solution remains practical, we rerank the documents based on the extracted rationales and fine-tune the reranker. This step distillates rationales from generators to rerankers, and aligns the reranker with the generator's information needs for answering the query effectively.

RADIO effectively addresses the preference inconsistency between RAG components by first generating a comprehensive rationale and then fine-tuning the reranker based on the extracted rationale. This approach considers the deeper reasoning behind answers. We evaluate RADIO on three tasks across four datasets: Open-domain QA (NQ (Kwiatkowski et al., 2019) and TriviaQA (Joshi et al., 2017)), Multi-choice questions (MMLU (Hendrycks et al., 2020)), and Multi-hop QA (Musique (Trivedi et al., 2022)). The results validate the superiority of our method compared to other state-of-the-art baselines. Our contributions can be summarized as follows:

- We propose RADIO, a novel and practical framework designed to address the preference misalignment of different components in RAG pipelines.
- We introduce rationale distillation within the RAG framework, which is an effective approach that leverages explicit textual rationales as signals to align the preferences of different components in RAG.
- Extensive experiments are conducted on three tasks across four datasets to demonstrate the effectiveness and transferability of RADIO.

2 Related Work

Retrieval-Augmented Generation. Large language models (LLMs) have demonstrated groundbreaking performance across numerous tasks (Jia et al., 2025) but still face challenges such as hallucination and outdated knowledge (Gao et al., 2023). To address these issues, retrieval-augmented generation (RAG) has been introduced (Fan et al., 2024; Zhang et al., 2025; Jia et al., 2024a). RAG retrieves relevant information from external knowledge bases and incorporates it as contextual input to the generator (LLM), enhancing the accuracy and reliability of the generated responses. The typical RAG pipeline can be divided into several key components: query rewriter, retriever, reranker, and generator. The query rewriter (Wang et al., 2023a) modifies and expands (Jia et al., 2024b; Li et al., 2023a) the original query to improve retrieval recall, ensuring more relevant documents are retrieved. The retriever (Chen et al., 2024) fetches relevant documents based on the query. Dense retrievers generally outperform sparse retrievers in this step. To integrate contextual information more effectively and identify documents more relevant to the query, rerankers (Moreira et al., 2024a) with larger models and greater complexity are introduced to reorder the retrieved documents compared to retrievers. Finally, the generator—usually a powerful LLM such as GPT-4 (Achiam et al., 2023) or Llama (Touvron et al., 2023)—uses the query and the top-k documents from the reranker to generate the final response. In this work, we address the issue of preference misalignment among different components within the RAG pipeline. We aim to leverage rationale as a signal to align these preferences and enhance the overall performance of the RAG system. It is worth noting that RADIO and methods like ReAct (Yao et al., 2022) are not mutually exclusive and can be used together to complement each other. While RADIO enhances the ranking of retrieved documents at each step, ReAct focuses on task-solving through reasoning and action.

Preference Alignment. To further improve LLMs, preference alignment is often performed after the initial pretraining phase (Jiang et al., 2024). Approaches such RLHF (Ouyang et al., 2022) and DPO (Rafailov et al., 2024) are proposed to align the output of LLMs more closely with human preferences. DPO transforms tasks into classification problems, achieving high computational efficiency and strong performance. In the context of RAG, several works (Ke et al., 2024) can be transformed to address the challenge of preference alignment between RAG components. REPLUG (Shi et al., 2023) improves RAG pipelines involving black-box LLMs by using the probability of the LLM generating the correct answer as a signal to determine document importance. Similarly, RRR (Ma et al., 2023) uses metrics based on the quality

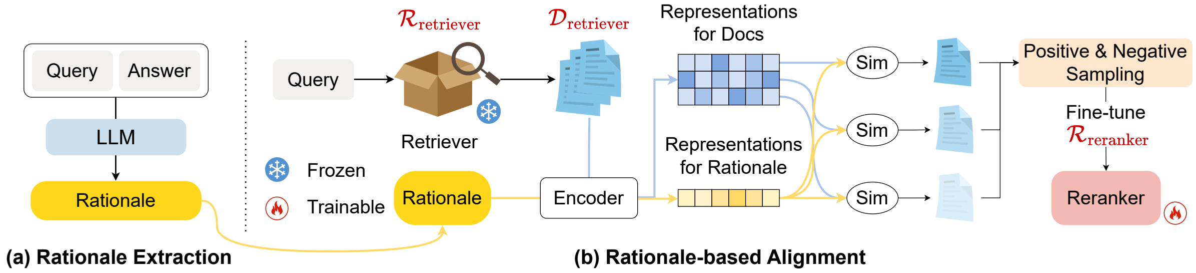

Query
Answer
LLM
Rationale
(a) Rationale Extraction
$\mathcal{R}_{\text{retriever}}$
$\mathcal{D}_{\text{retriever}}$
Query
Retriever
Frozen
Trainable
Rationale
Encoder
Representations for Docs
Representations for Rationale
Sim
Sim
Sim
Positive & Negative Sampling
Fine-tune $\mathcal{R}_{\text{reranker}}$
Reranker
(b) Rationale-based Alignment

Figure 1: Overview of RADIO.

of the LLM's generated response as a signal to evaluate a document's utility. On the other hand, ARL2 (Zhang et al., 2024) prompts LLMs to generate self-guided relevance labels for fine-tuning retriever, and DPA-RAG (Dong et al., 2024) introduces a bidirectional alignment strategy to mitigate preference inconsistencies in RAG pipelines. In this work, we focus on optimizing the reranker within RAG. Our goal is to enable the reranker to effectively identify supportive documents well-suited for the generator, whether black-box or open-source, facilitating the production of accurate outputs. In addition, BGM (Ke et al., 2024) trains a bridge model between retriever and LLMs to transform the retrieved information into the format LLM's prefer. Our method is theoretically compatible with BGM and could be combined to jointly enhance performance without any conflict.

3 Methodology

In this section, we detail rationale distillation in RAG. Specifically, we first demonstrate the task definition of RAG in Section 3.1. Then we give an overview of our proposed framework in Section 3.2, introduce the rationale extraction method in Section 3.3, and detail the rationale-based alignment in Section 3.4 and optimization in Section 3.5.

3.1 Task Definition

To address the hallucination problem and enhance adaptability to dynamic information of LLMs, RAG systems have been proposed. These systems enhance generative models by introducing additional contextual information retrieved based on a given query q. Specifically, when a query q is input into the RAG pipeline, the retriever $\mathcal{R}_{\text{retriever}}$ first retrieves relevant documents by calculating similarity scores and top-k_1 selection. The process can be formalized as follows:

$$\mathcal{D}_{\text{retriever}} = \{\boldsymbol{d}_i \mid \boldsymbol{d}_i \in \text{Top-}k_1(\text{score}_{\text{retriever}}(q, d))\} \tag{1}$$

where q and d are the query and document, $\text{score}_{\text{retriever}}$ denotes the score function in retriever, $\boldsymbol{d}_i$ means the i-th document in corpus and $\mathcal{D}_{\text{retriever}}$ is the documents set output by retriever $\mathcal{R}_{\text{retriever}}$.

To eliminate contextually irrelevant noise and provide more precise contextual information, the initially filtered documents will be further reranked by the reranker:

$$\mathcal{D}_{\text{reranker}} = \{\boldsymbol{d}_j \mid \boldsymbol{d}_j \in \text{Top-}k_2(\text{score}_{\text{reranker}}(q, d))\} \tag{2}$$

where $\mathcal{D}_{\text{reranker}}$ is the documents selected by reranker, $\text{score}_{\text{reranker}}$ denotes the score function in reranker. Note that the jth document $\boldsymbol{d}_j$ is in $\mathcal{D}_{\text{retriever}}$ (i.e., $d_j \in \mathcal{D}_{\text{retriever}}$) and k_2 is the number of documents selected by reranker, which is smaller than k_1 used by the retriever.

Finally, the documents filtered by the reranker, along with the original query, will be fed into the generator as contextual information to help generate the final response:

$$\hat{y} = \mathcal{G}(q, \mathcal{D}_{\text{reranker}}) \tag{3}$$

where $\hat{y}$ is the generated response and $\mathcal{G}$ denotes the generator.

It is worth noting that the documents selected by the reranker directly influence the generator's input. Therefore, in this work, we aim to align the preferences of the reranker and generator to enhance their consistency. This alignment improves the overall accuracy of the RAG system's responses.

3.2 Framework Overview

The overview of RADIO is depicted in Figure 1. RADIO is consisted of two phases: rationale extraction (Figure 1(a)) and rationale-based alignment (Figure 1(b)).

In the rationale extraction process, we combine the query with its ground truth answer and input them into LLMs to generate precise rationales. Using the correct answer as context in the prompt improves the accuracy of the LLM's rationale generation, ensuring that the generated rationale closely aligns with the requirements for deriving the correct answer.

In the rationale-based alignment process, our goal is to use the generated rationale to guide rerankers, enabling it to select documents that better support the generator in answering the query. Specifically, we leverage the generated rationale as a signal to rerank documents. The reranked documents will be used to fine-tune the reranker, addressing the preference misalignment between the reranker and the generator. By aligning these components, the process ensures that the selected documents are not only contextually relevant but also optimally supportive for the generator's reasoning and response generation.

3.3 Rationale Extraction

Rationales are critical components of LLM reasoning processes and have been shown to enhance the accuracy of LLM-generated responses significantly. This perspective is supported by existing works such as Chain-of-Thoughts (CoT (Wei et al., 2022)). Existing work (Shi et al., 2023; Ma et al., 2023; Dong et al., 2024) has primarily focused on the initial relationships between queries and documents or indirect relationships between answers and documents, while overlooking rationales, a crucial intermediary component in the reasoning process. Motivated by this, we aim to extract rationales as signals to align the preferences of different components in the RAG pipeline.

To accurately extract the rationale necessary for answering a query and deriving the correct answer, we combine the query with the ground truth answer as contextual information of LLMs, as shown in Figure 1(a). The prompt template used for this process is as follows: "*You are a professional QA assistant. Given a question and the ground truth answer, you can output the rationale why the ground truth answer is correct. Question: {question}. Answer: {answer}. Rationale:* ". The generation process can be formalized as:

$$r = \text{LLM}(q, a) \tag{4}$$

where q and a are the query and answer, r deontes the generated rationale.

By doing so, we effectively bridge the gap between the query and the answer by generating the necessary rationale. This rationale accurately supports the reasoning process required to derive the correct answer from the query.

3.4 Rationale-based Alignment

Given the extracted rationale r, a key challenge lies in effectively and efficiently utilizing it to improve preference consistency within the RAG pipeline. In this section, we propose a rationale-based alignment approach, where the rationale serves as a signal to fine-tune the reranker. This enables the reranker to identify and prioritize supportive documents that facilitate the generator in producing accurate responses. Specifically, we first use the retriever $\mathcal{R}_{\text{retriever}}$ to retrieve k_1 relevant documents based on the query:

$$\mathcal{D}_{\text{retriever}} = \mathcal{R}_{\text{retriever}}(q, \mathcal{C}) \tag{5}$$

where $\mathcal{D}_{retriever}$ is the document set retrieved by $\mathcal{R}_{\text{retriever}}$ based on query q and corpus $\mathcal{C}$, and $|\mathcal{D}_{retriever}| = k_1$.

Next, to facilitate the comparison of similarity between different documents and the rationale, we use a text encoder to convert both the documents and the rationale into dense vectors.

$$\boldsymbol{e}_i^{\text{document}} = \text{Encoder}(\boldsymbol{d}_i) \tag{6}$$

$$\boldsymbol{e}^{\text{rationale}} = \text{Encoder}(r) \tag{7}$$

where $\boldsymbol{e}_i^{\text{document}}$ and $\boldsymbol{e}^{\text{rationale}}$ denote the representations of ith document and rationale. $\boldsymbol{d}_i$ is the ith document and r represents the extracted rationale.

Then, we calculate the semantic similarity between each document and the rationale. Here, we use cosine similarity, denoted as $\text{sim}(\cdot)$. The calculated scores indicate the degree to which each document supports generating the correct answer, with higher scores reflecting stronger support. We also linearly interpolate the score of documents with their retrieval score in the retrieval stage by weighted score sum.

$$\boldsymbol{s}_i^{\text{rationale}} = \text{sim}(\boldsymbol{e}_i^{\text{document}}, \boldsymbol{e}^{\text{rationale}}) \tag{8}$$

$$\boldsymbol{s}_i^{\text{retriever}} = \text{score}_{\text{retriever}}(q, \boldsymbol{d}_i) \tag{9}$$

$$\boldsymbol{s}_i' = \alpha \boldsymbol{s}_i^{\text{rationale}} + (1 - \alpha)\boldsymbol{s}_i^{\text{retriever}} \tag{10}$$

where $\boldsymbol{s}_i^{\text{rationale}}$, $\boldsymbol{s}_i^{\text{retriever}}$, $\boldsymbol{s}_i'$ represent the rationale similarity score, retrieval score, and final score for ith document. $\text{score}_{\text{retriever}}(\cdot)$ is the score function in retriever and α is a hyperparameter used for integration. Note that we apply min-max normalization in this work to both rationale score and retrieval score before integration.

Next, we rerank the documents based on their scores. Following the previous sampling method *Top-k shifted by N* (Moreira et al., 2024b), we select the top-ranked document as the positive sample and then shift by n documents and sample m negative samples from the subsequent documents to construct positive-negative pairs for fine-tuning the reranker. This process can be represented as:

$$\boldsymbol{d}_{\text{pos}} = \boldsymbol{d}_i, \text{where } i = \arg\max_i \boldsymbol{s}'_i \tag{11}$$

$$\{\boldsymbol{d}_{\text{neg}}\} = \text{Sample}_m(\{\boldsymbol{d}_i | \text{rank}(\boldsymbol{d}_i) > n\}) \tag{12}$$

where $\boldsymbol{d}_{\text{pos}}$ and $\boldsymbol{d}_{\text{neg}}$ are the sampling positive and negative documents, $\text{Sample}_m(\cdot)$ denotes a sampling operation that selects m negative documents from the set of documents ranked lower than n.

3.5 Optimization

Following BGE embedding (Xiao et al., 2023) and QA Ranking Benchmark (Moreira et al., 2024a), we use InfoNCE as our optimization objectives to fine-tune reranker:

$$f(q, d) = \exp(\phi(q, d)/\tau) \tag{13}$$

$$L = -\log \frac{f(q, d^+)}{f(q, d^+) + \sum_{i=1}^{N} f(q, d_i^-)} \tag{14}$$

where d^+ and d^- represent the positive and negative document, τ is the temperature parameter, and N denotes the number of negative documents.

4 Experiments

4.1 Datasets and Metrics

Following previous work (Shi et al., 2023; Zhang et al., 2024), we conduct experiments on three tasks across four datasets to evaluate RADIO with other methods: **Open-domain QA** (NQ (Kwiatkowski et al., 2019) and TriviaQA (Joshi et al., 2017)), **Multi-choice questions** (MMLU (Hendrycks et al., 2020)), and **Multi-hop QA** (Musique (Trivedi et al., 2022)). Detailed dataset descriptions are given in Appendix A.1. Following previous work (Ma et al., 2023; Shi et al., 2023), we report EM and F1 scores for QA datasets and EM for MMLU.

4.2 Baselines

To verify the effectiveness of RADIO, we conduct experiments with the following baseline methods: **Base** (Xiao et al., 2023), **Atlas** (Izacard et al., 2023), **REPLUG** (Shi et al., 2023), **Trainable rewrite-retrieve-read (RRR)** (Ma et al., 2023), **ARL2** (Zhang et al., 2024), and **DPA-RAG** (Dong et al., 2024). The detailed introduction of baselines is given in Appendix A.2.

4.3 Backbone Rerankers

To validate the generality and adaptability of RADIO, we select three different rerankers as the backbone models for our experiments: **gte-base** (Li et al., 2023b), **gte-large** (Li et al., 2023b), and **bge-reranker-base** (Xiao et al., 2023). A detailed introduction is given in Appendix A.3.

4.4 Implementation Details

We implement RADIO on FlashRAG (Jin et al., 2024), a Python library for efficient RAG research. In the RAG pipeline, we take *e5-base-v2* (Wang et al., 2022) as the retriever, and *Meta-Llama-3.1-8B-Instruct* (Touvron et al., 2023) as the generator. We sample 20,000 instances from NQ and TriviaQA to construct the fine-tuning dataset and fine-tune rerankers separately. For document sampling, we set the shift n in *Top-k shifted by N* method as 3, and sample 6 negative samples from the subsequent documents. To ensure a fair comparison, the sampling index is fixed and remains unchanged across methods. In the RAG pipeline, we set the number of documents selected by retriever and reranker (i.e., k_1 and k_2) as 20 and 5. The retrieval corpus we used in experiments is Wikipedia (2018, December). For fine-tuning the reranker, we tune the training epochs from 1 to 5 and the integration hyperparameter α from 0.0 to 1.0. We use Adam (Kingma, 2014) optimizer with a learning rate 6e-5 and a weight decay of 0.01. The prompts we used in experiments are given in Appendix A.4. We provide an analysis of the impact of fine-tuning sample size on performance in Appendix A.5, and we also give a detailed analysis on the shift parameter n and the negative samples m in Appendix A.6.

4.5 Main Results

4.5.1 Open-domain QA

To evaluate the effectiveness of RADIO and its transferability across different rerankers, we conduct experiments on the NQ and TriviaQA datasets. The results are presented in Table 1. A case study is given in Appendix A.7. From these results, we can draw the following conclusions:

- Compared to the Base method, most experimen-

Table 1: Overall experiments. “*” indicates the statistically significant improvements (i.e., two-sided t-test with $p < 0.05$) over the best baseline. For all metrics, higher is better. Δ represents the relative improvement of RADIO over Base method.

Method	NQ						TriviaQA					
	gte-base		gte-large		bge-reranker-base		gte-base		gte-large		bge-reranker-base	
	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1	EM	F1
Base	0.2931	0.4046	0.2798	0.3935	0.3371	0.4603	0.5449	0.6374	0.5495	0.6414	0.6114	0.7120
Atlas	0.3338	0.4587	0.3418	0.4677	0.3521	0.4832	0.5823	0.6752	0.6004	0.6972	0.6083	0.7063
REPLUG	0.3257	0.4484	0.2607	0.3670	0.3427	0.4753	0.5679	0.6578	0.5248	0.6171	0.6032	0.7004
RRR	0.3374	0.4608	0.3299	0.4578	0.3438	0.4754	0.5801	0.6716	0.5358	0.6237	0.6099	0.7091
ARL2	0.3413	0.4688	0.3515	0.4804	0.3568	0.4885	0.6079	0.7086	0.6107	0.7120	0.6137	0.7149
DPA-RAG	0.3391	0.4674	0.3385	0.4710	0.3462	0.4793	0.6080	0.7076	0.6097	0.7119	0.6149	**0.7169***
RADIO (Ours)	**0.3512***	**0.4790***	**0.3565***	**0.4850***	**0.3665***	**0.4917***	**0.6084**	**0.7095***	**0.6128***	**0.7137***	**0.6154**	0.7151
Δ	↑ 19.82%	↑ 18.39%	↑27.41%	↑23.25%	↑8.72%	↑6.82%	↑11.55%	↑11.31%	↑11.52%	↑11.27%	↑0.65%	↑0.40%

Table 2: Experimental results on MMLU. EM is reported as the evaluation metric. The source dataset used to fine-tune rerankers is the Open-domain QA dataset NQ. Δ represents the relative improvement of RADIO over Base method.

Method	Humanities	Social	STEM	Other	ALL
Base	0.4089	0.6867	**0.5147**	0.6650	0.5502
Atlas	0.3985	0.6935	0.5074	0.6563	0.5447
REPLUG	0.4102	0.6854	0.5065	0.6590	0.5473
RRR	0.4079	0.6913	0.5116	0.6572	0.5484
ARL2	0.4147	0.7016	0.5106	0.6630	0.5540
DPA-RAG	0.4157	0.701	0.5078	0.6652	0.5541
RADIO (Ours)	**0.4172**	**0.7013**	0.5080	**0.6717**	**0.5562**
Δ	↑2.03%	↑2.13%	↓1.30%	↑1.01%	↑1.09%

tal settings achieve better results, demonstrating the necessity of preference alignment within the RAG pipeline.

- Compared to other baseline methods, RADIO consistently achieves superior performance across all datasets and reranker backbone configurations. The results validate the effectiveness of using rationales as signals for preference alignment in RAG pipeline.

- On TriviaQA, methods such as RRR and REPLUG show performance declines relative to the base method when using rerankers *gte-large* and *bge-reranker-base*. This indicates that these methods are sensitive, limiting their applicability. In contrast, RADIO demonstrates robust adaptability to different rerankers, achieving significant performance improvements across all three rerankers.

- As the reranker becomes larger or more powerful (e.g., progressing from *gte-base* to *gte-large* and further to *bge-reranker-base*), the performance ranking of models fine-tuned with RADIO aligns with the reranker's inherent capabilities. This suggests that RADIO's performance gains are sustainable and scalable with stronger rerankers, providing an avenue to further explore the upper performance limits of RAG pipelines.

4.6 Task Generalization Analysis

4.6.1 Multi-choice Questions

We also conduct experiments on MMLU. Since MMLU is a multiple-choice dataset, we report the EM metric (Ma et al., 2023). Additionally, following previous work (Yu et al., 2023), we fine-tune the reranker *bge-reranker-base* using open-domain QA as the source task and evaluate its performance on the MMLU dataset. Table 2 shows the results of fine-tuning reranker with NQ dataset. The results of fine-tuning reranker with TriviaQA are given in Appendix A.8. We can draw the following conclusions:

- From the metrics corresponding to the ALL category, RADIO demonstrates consistent improvements over Base. This highlights the effectiveness and generalization of RADIO, as it successfully adapts to multi-choice question tasks even when fine-tuned on the Open-domain QA tasks.

- Analyzing the results by question category, RADIO shows more significant improvements over the Base method in the Humanities and Social Sciences categories, with average gains of **2.03%** and **2.13%**, respectively. However, it exhibits a slight negative effect in the STEM category. This may be due to the fine-tuning datasets (NQ and TriviaQA), which are Open-domain QA datasets with distributions more similar to humanities and social sciences but markedly different from STEM subjects.

- Compared to other baseline methods, RADIO achieves top performance in the vast majority of

Table 3: Transferability analysis across generators. EM is reported as the metric for MMLU dataset.

Generator	Method	NQ		MMLU				
		EM	F1	Humanities	Social	STEM	Other	ALL
Llama3.1-8b-instruct	Base	0.3371	0.4603	0.4089	0.6867	**0.5147**	0.6650	0.5502
	RADIO (ours)	**0.3665**	**0.4917**	**0.4172**	**0.7013**	0.5080	**0.6717**	**0.5562**
qwen2.5-14b-instruct	Base	0.3310	0.4484	0.5439	0.8200	0.7206	0.7541	0.6906
	RADIO (ours)	**0.3518**	**0.4753**	**0.5598**	**0.8229**	**0.7250**	**0.7631**	**0.6995**
gpt4o-mini	Base	0.3607	0.4880	0.6485	0.8362	**0.6784**	0.8005	0.7300
	RADIO (ours)	**0.3742**	**0.5086**	**0.6548**	**0.8372**	0.6768	**0.8010**	**0.7321**

Table 4: Experimental results on Musique. EM and F1 are reported as the metrics. The source datasets used to fine-tune rerankers are Open-domain QA datasets NQ and TriviaQA. Δ represents the relative improvement of RADIO over Base method.

Method	NQ (Source Dataset)		TriviaQA (Source Dataset)	
	EM	F1	EM	F1
Base	0.0662	0.1297	0.0662	0.1297
Altlas	0.0658	0.1336	0.0541	0.1184
REPLUG	0.0608	0.1269	0.0559	0.1188
RRR	0.0658	0.1315	0.0592	0.1207
ARL2	0.0641	0.1309	0.0674	0.1330
DPA-RAG	0.0629	0.1307	0.0690	0.1337
RADIO (Ours)	**0.0699**	**0.1371**	**0.0695**	**0.1347**
Δ	↑5.59%	↑5.71%	↑4.98%	↑3.86%

metrics, demonstrating its superiority and state-of-the-art capability.

4.6.2 Multi-hop QA

We also conduct experiments on the multi-hop dataset Musique with reranker *bge-reranker-base*, and the experimental results are shown in Table 4. Based on the experimental results, we can draw the following conclusions:

- First, RADIO achieves significant performance improvements in all evaluation metrics (EM and F1), demonstrating its task generalization ability in multi-hop reasoning tasks. On the NQ source dataset, RADIO achieves an EM value of 0.0699 and an F1 value of 0.1371, with improvements of 5.59% and 5.71% over the Base method, respectively. On the TriviaQA source dataset, RADIO's EM value is 0.0695, and F1 value is 0.1347, showing improvements of 4.98% and 3.86% over the Base method.

- Secondly, compared to other baseline methods, RADIO demonstrates a more balanced and consistent advantage across both the NQ and TriviaQA source datasets. Notably, when using NQ as the source dataset, only RADIO shows a positive improvement in the EM metric, while all other baseline methods experience varying degrees of negative degradation.

4.7 Transferability Analysis across Generators

We conduct experiments on two datasets of different tasks, NQ and MMLU, using three different generators (*Llama3.1-8b-instruct* (Touvron et al., 2023), *qwen2.5-14b-instruct* (Yang et al., 2024), and *gpt4o-mini* (Achiam et al., 2023)) with *bge-reranker-base* fine-tuned on NQ to validate the transferability of our method, as shown in Table 3. We can find that (1) RADIO maintains its effectiveness across different generators, consistently enhancing the performance of the original RAG pipeline. This demonstrates RADIO's strong transferability with various generators. (2) Comparing different generators reveals that RADIO's performance gains are more pronounced with smaller, less capable generators. Specifically, when the generators are Llama3.1, Qwen2.5, and GPT4o-mini, RADIO achieves EM improvements of 8.72%, 6.28%, and 3.74%, respectively, and F1 improvements of 6.82%, 6.00%, and 4.22%. This is because as the generator's capability increases and approaches the upper performance limits of the RAG pipeline, further enhancing the pipeline becomes increasingly challenging, resulting in a smaller improvement.

4.8 Ablation Study

To explore the specific impact of rationale and retrieval score, we design the following variants with reranker *bge-reranker-base*:

- **w/o ALL:** Base reranker without fine-tuning. Do not introduce rationale or retrieval score.

- **w/o Retrieval:** Ranking documents and fine-tuning reranker only based on the rationale scores.

Table 5: Ablation study.

Dataset	Metrics	w/o ALL	w/o Retrieval	RADIO
NQ	EM	0.3371	0.3587	**0.3665**
	F1	0.4603	0.4858	**0.4917**
MMLU	Humanities (EM)	0.4089	0.4168	**0.4172**
	Social (EM)	0.6867	0.6981	**0.7013**
	STEM (EM)	**0.5147**	0.5109	0.508
	Other (EM)	0.665	0.6666	**0.6717**
	ALL (EM)	0.5502	0.5548	**0.5562**

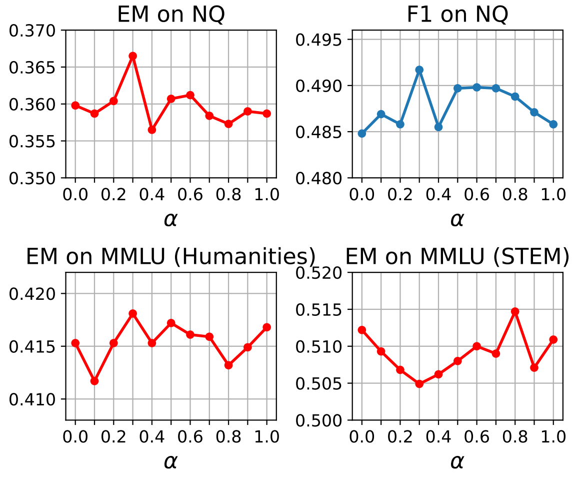


Figure 2: Hyperparameter analysis on NQ and MMLU.

- **RADIO:** Fine-tuning reranker based on both rationale and retrieval scores.

Table 5 shows the results of ablation study on NQ and MMLU, where we can derive the following findings: (1) Both the rationale score and retrieval score contribute positively to RADIO's performance, with the rationale score demonstrating a stronger positive impact compared to the retrieval score. (2) RADIO outperforms both **w/o ALL** and **w/o Retrieval**, while **w/o Retrieval** surpasses **w/o ALL**. This indicates that the rationale score and retrieval score are not conflicting but rather complementary. Their integration provides a more robust signal for document ranking, which effectively aids in fine-tuning the reranker. (3) In the STEM category of the MMLU dataset, **w/o ALL** outperforms both **w/o Retrieval** and RADIO. This could be attributed to the fact that the training dataset (NQ) contains questions with distributions more similar to humanities and social sciences, leading to trends in the Humanities, Social Sciences, and Other categories that differ from STEM category.

4.9 Hyperparameter Analysis

Figure 2 visualizes the performance of RADIO across different integration coefficients α on NQ and MMLU. The x-axis represents the integration coefficients α and the y-axis represents the evaluation metrics EM (red) or F1 (blue). For the MMLU dataset, we present results for two representative categories: Humanities and STEM. The trends for other categories align with those observed in Humanities. Complete experimental results are provided in Appendix A.9 for reference. From the figure, we can draw the following conclusions: (1) As α increases, the metrics on the NQ dataset and most categories of the MMLU dataset exhibit a trend of first rising and then falling, with the optimal range for α being between 0.3 and 0.7. This demonstrates the complementary nature of rationale scores and retrieval scores, which together form an optimal signal for preference alignment. (2) When $\alpha = 0$, the RAG performance is suboptimal because document scoring relies entirely on retrieval scores, focusing solely on query-document relevance while ignoring whether the document supports the generator in answering the query. Conversely, when $\alpha = 1$, the performance is still not optimal, as it completely disregards retrieval relevance, leading to a mismatch between the fine-tuning dataset and the training data, which negatively affects model performance. (3) The trends for STEM differ from those of other MMLU categories, showing an opposite pattern. This is likely due to the significant distributional differences between STEM and Humanities/Social Sciences, resulting in a "seesaw effect" as observed in the figure. It is worth noting that this phenomenon is also reflected in REPLUG (Shi et al., 2023), where the improvement in the STEM category is weaker compared to other categories.

5 Conclusion

In this paper, we propose a novel and practical preference alignment framework, RADIO, with rationale distillation in retrieval-augmented generation. First, we introduce a rationale extraction method to extract the rationales necessary for answering queries with LLMs. Next, a rationale-based alignment is proposed to rerank documents based on extracted rationales and fine-tune rerankers. Extensive experiments on three tasks across four datasets are conducted to validate the effectiveness of our proposed method against state-of-the-art baselines and demonstrate its strong transferability. We also release our code online to facilitate the following research and ease of reproduction.

6 Limitations

First, compared to other methods, our approach RADIO requires additional time in the rationale extraction stage to generate rationales. Since different samples are independent of one another, we can reduce generation time by employing parallel processing to mitigate this issue. Secondly, the MMLU experimental results reveal that the composition of fine-tuning datasets can affect RADIO's effectiveness. This issue can be addressed by designing task-specific fine-tuning datasets for different downstream tasks or large-scale general fine-tuning datasets.

Acknowledgments

This research was partially supported by Research Impact Fund (No.R1015-23), Collaborative Research Fund (No.C1043-24GF), and Huawei (Huawei Innovation Research Program, Huawei Fellowship).

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

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.

Guanting Dong, Yutao Zhu, Chenghao Zhang, Zechen Wang, Zhicheng Dou, and Ji-Rong Wen. 2024. Understand what llm needs: Dual preference alignment for retrieval-augmented generation. *arXiv preprint arXiv:2406.18676*.

Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li. 2024. A survey on rag meeting llms: Towards retrieval-augmented large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6491–6501.

Zichuan Fu, Xiangyang Li, Chuhan Wu, Yichao Wang, Kuicai Dong, Xiangyu Zhao, Mengchen Zhao, Huifeng Guo, and Ruiming Tang. 2023. A unified framework for multi-domain ctr prediction via large language models. *ACM Transactions on Information Systems*.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

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

Gautier Izacard, Patrick Lewis, Maria Lomeli, Lucas Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-Yu, Armand Joulin, Sebastian Riedel, and Edouard Grave. 2023. Atlas: Few-shot learning with retrieval augmented language models. *Journal of Machine Learning Research*, 24(251):1–43.

Pengyue Jia, Qingpeng Cai, Xiangyu Zhao, Ling Pan, Xin Xin, Jin Huang, Weinan Zhang, Li Zhao, Dawei Yin, and Grace Hui Yang. 2025. Agentir: 2nd workshop on agent-based information retrieval. ACM SIGIR.

Pengyue Jia, Yiding Liu, Xiaopeng Li, Xiangyu Zhao, Yuhao Wang, Yantong Du, Xiao Han, Xuetao Wei, Shuaiqiang Wang, and Dawei Yin. 2024a. G3: an effective and adaptive framework for worldwide geolocalization using large multi-modality models. *Advances in Neural Information Processing Systems*, 37:53198–53221.

Pengyue Jia, Yiding Liu, Xiangyu Zhao, Xiaopeng Li, Changying Hao, Shuaiqiang Wang, and Dawei Yin. 2024b. Mill: Mutual verification with large language models for zero-shot query expansion. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2498–2518.

Ruili Jiang, Kehai Chen, Xuefeng Bai, Zhixuan He, Juntao Li, Muyun Yang, Tiejun Zhao, Liqiang Nie, and Min Zhang. 2024. A survey on human preference learning for large language models. *arXiv preprint arXiv:2406.11191*.

Jiajie Jin, Yutao Zhu, Xinyu Yang, Chenghao Zhang, and Zhicheng Dou. 2024. Flashrag: A modular toolkit for efficient retrieval-augmented generation research. *arXiv preprint arXiv:2405.13576*.

Mandar Joshi, Eunsol Choi, Daniel S Weld, and Luke Zettlemoyer. 2017. Triviaqa: A large scale distantly supervised challenge dataset for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1601–1611.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Zixuan Ke, Weize Kong, Cheng Li, Mingyang Zhang, Qiaozhu Mei, and Michael Bendersky. 2024. Bridging the preference gap between retrievers and llms. *arXiv preprint arXiv:2401.06954*.

Diederik P Kingma. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. 2019. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, 7:453–466.

Xiaopeng Li, Lixin Su, Pengyue Jia, Xiangyu Zhao, Suqi Cheng, Junfeng Wang, and Dawei Yin. 2023a. Agent4ranking: Semantic robust ranking via personalized query rewriting using multi-agent llm. *arXiv preprint arXiv:2312.15450*.

Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. 2023b. Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*.

Qidong Liu, Xian Wu, Xiangyu Zhao, Yuanshao Zhu, Zijian Zhang, Feng Tian, and Yefeng Zheng. 2024. Large language model distilling medication recommendation model. *arXiv preprint arXiv:2402.02803*.

Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao, and Nan Duan. 2023. Query rewriting for retrieval-augmented large language models. *arXiv preprint arXiv:2305.14283*.

Gabriel de Souza P Moreira, Ronay Ak, Benedikt Schifferer, Mengyao Xu, Radek Osmulski, and Even Oldridge. 2024a. Enhancing q&a text retrieval with ranking models: Benchmarking, fine-tuning and deploying rerankers for rag. *arXiv preprint arXiv:2409.07691*.

Gabriel de Souza P Moreira, Radek Osmulski, Mengyao Xu, Ronay Ak, Benedikt Schifferer, and Even Oldridge. 2024b. Nv-retriever: Improving text embedding models with effective hard-negative mining. *arXiv preprint arXiv:2407.15831*.

Wei Ni, Kaihang Zhang, Xiaoye Miao, Xiangyu Zhao, Yangyang Wu, Yaoshu Wang, and Jianwei Yin. 2025. Zeroed: Hybrid zero-shot error detection through large language model reasoning. *arXiv preprint arXiv:2504.05345*.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Weijia Shi, Sewon Min, Michihiro Yasunaga, Minjoon Seo, Rich James, Mike Lewis, Luke Zettlemoyer, and Wen-tau Yih. 2023. Replug: Retrieval-augmented black-box language models. *arXiv preprint arXiv:2301.12652*.

Devendra Singh, Siva Reddy, Will Hamilton, Chris Dyer, and Dani Yogatama. 2021. End-to-end training of multi-document reader and retriever for open-domain question answering. *Advances in Neural Information Processing Systems*, 34:25968–25981.

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

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2022. Musique: Multi-hop questions via single-hop question composition. *Transactions of the Association for Computational Linguistics*, 10:539–554.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. 2022. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*.

Liang Wang, Nan Yang, and Furu Wei. 2023a. Query2doc: Query expansion with large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9414–9423.

Maolin Wang, Yao Zhao, Jiajia Liu, Jingdong Chen, Chenyi Zhuang, Jinjie Gu, Ruocheng Guo, and Xiangyu Zhao. 2023b. Large multimodal model compression via efficient pruning and distillation at antgroup. *arXiv preprint arXiv:2312.05795*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-pack: Packaged resources to advance general chinese embedding. *Preprint*, arXiv:2309.07597.

Derong Xu, Pengyue Jia, Xiaopeng Li, Yingyi Zhang, Maolin Wang, Qidong Liu, Xiangyu Zhao, Yichao Wang, Huifeng Guo, Ruiming Tang, et al. 2025a. Align-grag: Reasoning-guided dual alignment for graph retrieval-augmented generation. *arXiv preprint arXiv:2505.16237*.

Derong Xu, Xinhang Li, Ziheng Zhang, Zhenxi Lin, Zhihong Zhu, Zhi Zheng, Xian Wu, Xiangyu Zhao, Tong Xu, and Enhong Chen. 2025b. Harnessing large language models for knowledge graph question answering via adaptive multi-aspect retrieval-augmentation.

In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 25570–25578.

Derong Xu, Ziheng Zhang, Zhenxi Lin, Xian Wu, Zhihong Zhu, Tong Xu, Xiangyu Zhao, Yefeng Zheng, and Enhong Chen. 2024. Multi-perspective improvement of knowledge graph completion with large language models. *arXiv preprint arXiv:2403.01972*.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zhihao Fan. 2024. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Zichun Yu, Chenyan Xiong, Shi Yu, and Zhiyuan Liu. 2023. Augmentation-adapted retriever improves generalization of language models as generic plug-in. *arXiv preprint arXiv:2305.17331*.

Lingxi Zhang, Yue Yu, Kuan Wang, and Chao Zhang. 2024. Arl2: Aligning retrievers for black-box large language models via self-guided adaptive relevance labeling. *arXiv preprint arXiv:2402.13542*.

Yingyi Zhang, Pengyue Jia, Xianneng Li, Derong Xu, Maolin Wang, Yichao Wang, Zhaocheng Du, Huifeng Guo, Yong Liu, Ruiming Tang, et al. 2025. Lsrp: A leader-subordinate retrieval framework for privacy-preserving cloud-device collaboration. *arXiv preprint arXiv:2505.05031*.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223*.

Xiangyu Zhao, Long Xia, Liang Zhang, Zhuoye Ding, Dawei Yin, and Jiliang Tang. 2018a. Deep reinforcement learning for page-wise recommendations. In *Proceedings of the 12th ACM conference on recommender systems*, pages 95–103.

Xiangyu Zhao, Liang Zhang, Zhuoye Ding, Long Xia, Jiliang Tang, and Dawei Yin. 2018b. Recommendations with negative feedback via pairwise deep reinforcement learning. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1040–1048.

A Appendix

A.1 Dataset Desctiptions

The detailed descriptions of baselines are given as follows:

- **Natural Questions (NQ):** NQ contains real user questions compiled from Google search, with corresponding answers identified from Wikipedia by human annotators.
- **TriviaQA:** TriviaQA dataset comprises trivia questions paired with answer annotations and supporting evidence documents, such as web pages and Wikipedia articles. It is designed to assess a model's ability to retrieve and comprehend textual evidence for open-domain question answering.
- **Massive Multitask Language Understanding (MMLU):** MMLU is a comprehensive evaluation dataset comprising 57 categories of questions, which are grouped into four broad domains: Humanities, Social Sciences, STEM, and Other. In this paper, we report evaluation metrics based on these categories.
- **Musique:** MuSiQue is a multi-hop question answering dataset created using a bottom-up approach to ensure connected reasoning, with 25,000 2–4 hop questions.

A.2 Baselines

Following is the introduction of baselines:

- **Base (Xiao et al., 2023):** The reranker model is used off-the-shelf without any fine-tuning.
- **Atlas (Izacard et al., 2023):** A pretrained retrieval-augmented language model designed for knowledge intensive task. We choose the EMDR^2 (Singh et al., 2021) as the reward to rerank documents and fine-tune rerankers.
- **REPLUG (Shi et al., 2023):** REPLUG seeks to fine-tune the retriever to enhance RAG pipelines that include black-box LLMs. It achieves this by using the query and document as contextual inputs and leveraging the probability of the LLM generating the correct answer as the importance score. This idea is also reflected in the PDist method in Atlas (Izacard et al., 2023).

Table 6: Prompts for QA datasets.

System Prompt: Answer the question based on the given document. Only give me the answer and do not output any other words. The following are given documents.
{reference}
User Prompt:
Question: {question}
Answer:

- **Trainable rewrite-retrieve-read (RRR) (Ma et al., 2023):** RRR optimizes the query rewriter using the evaluation metrics of the final RAG output as a reward, which is used to fine-tune rerankers in our pipeline, enhancing the overall effectiveness of RAG.
- **ARL2 (Zhang et al., 2024):** ARL2 introduces a method to use LLMs as supervisor to generate self-guided relevance labels (e.g., "Not relevant", "relevant") for fine-tuning retriever.
- **DPA-RAG (Dong et al., 2024):** DPA-RAG proposes a knowledge preference pipeline to dual-align rerankers and generators in RAG. It combines document importance from the LLM's perspective with the importance determined during the retrieval stage.

A.3 Detailed Introduction to Backbone Rerankers

To validate the generality and adaptability of RADIO, we select three different rerankers as the backbone models for our experiments: (1) **gte-base** (Li et al., 2023b): a reranker model proposed by Alibaba DAMO Academy, with 109M parameters and 768 embedding dimensions. (2) **gte-large** (Li et al., 2023b): the larger version of gte-base, with 335M parameters and 1024 embedding dimensions. (3) **bge-reranker-base** (Xiao et al., 2023): a powerful cross-encoder architecture reranker proposed by Beijing Academy of Artificial Intelligence, with 278M parameters.

A.4 Prompts

In this section, we detail the prompts we used in the experiments. For Open-domain QA datasets (NQ and TriviaQA) and Multi-hop QA dataset (Musique), the prompts are as shown in Table 6. For Multi-choice dataset MMLU, the prompts are shown in Table 7.

Table 7: Prompts for Multi-choice datasets.

System Prompt: Answer the question based on the given document. Only give me the option (A/B/C/D) and do not output any other words. The following are given documents.
{reference}
User Prompt:
Question: {question}
Answer:

Table 8: Experimental results on NQ with different numbers of fine-tuning samples.

Sample Num.	EM	F1
0	0.3371	0.4603
4,000	0.3526	0.4843
8,000	0.3634	0.4896
12,000	0.3582	0.4842
16,000	0.3604	0.4889
20,000	0.3665	0.4917

A.5 Sample Efficiency Analysis of Fine-tuning

To understand the impact of fine-tuning samples on RAG performance, we conduct additional experiments to evaluate RADIO's performance under a few-shot learning setup using varying amounts of fine-tuning samples (0, 4,000, 8,000, 12,000, 16,000, and 20,000 instances) on the NQ dataset with *bge-reranker-base*. The results are shown in Table 8.

We observe that even with a relatively small number of fine-tuning samples (e.g., 4,000 or 8,000), RADIO yields substantial improvements over the zero-shot baseline. Notably, with 8,000 samples the EM score increases by 7.8% and the F1 score by 6.4%, clearly demonstrating the model's capacity to benefit from limited task-specific data. Moreover, as the number of fine-tuning instances increases, the overall performance tends to improve, suggesting that additional annotated data can further enhance the retrieval and generation capabilities of the system. Although there are minor fluctuations—such as a slight dip in performance at 12,000 and 16,000 samples—these variations may stem from differences in data distribution or optimization dynamics.

Table 9: Hyperparameter analysis on the number of negative samples m.

Shift (n)	Negative Samples m	EM	F1
3	2	0.3615	0.4886
3	4	0.3601	0.4888
3	6	0.3665	0.4917
3	8	0.3604	0.4895

Table 10: Hyperparameter analysis on the shift n in *Top-k shifted by N*

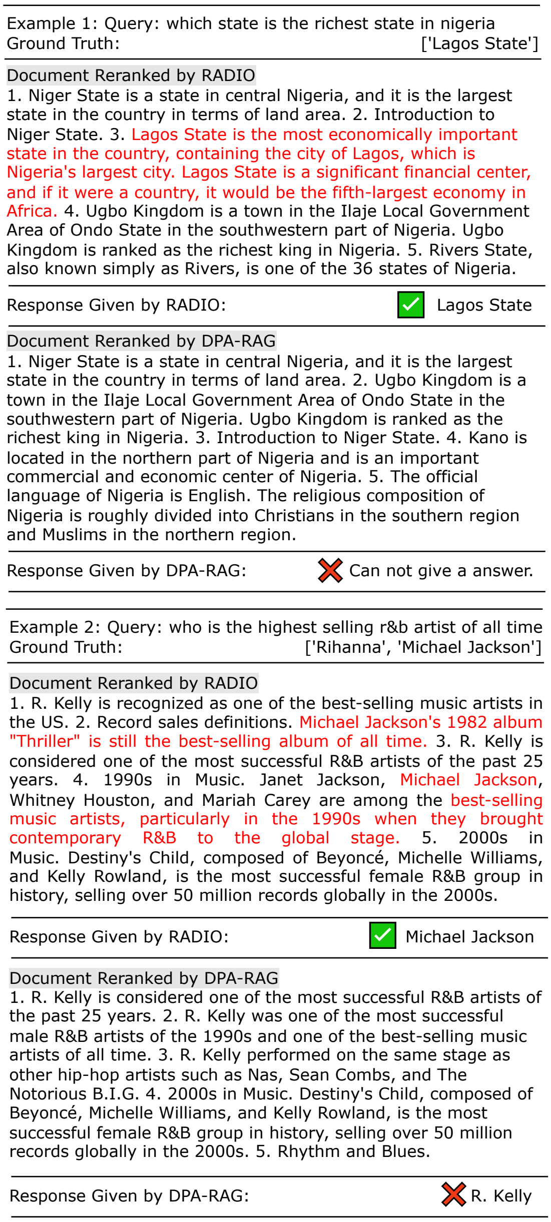
Shift (n)	Negative Samples m	EM	F1
1	6	0.3654	0.4918
2	6	0.3554	0.4826
3	6	0.3665	0.4917
4	6	0.3579	0.4868
5	6	0.3587	0.4887

A.6 More Analysis on Training Hyperparameters

To understand the specific impact of two training hyperparameters (i.e., the shift n in *Top-k shifted by N* and the number of negative samples m) on performance, we conduct the following experiments on NQ based on *bge-reranker-base*:

- **Varying** m: We fix $n = 3$ and vary m over the values 2,4,6, and 8.
- **Varying** n: We fix $m = 6$ and vary n over the values 1,2,3,4, and 5.

The detailed experimental results are presented in Table 9 and Table 10, respectively. From the results, we can find the following points: (1) **Impact of Negative Samples** m: Variations in the number of negative samples m have a relatively modest effect on RADIO's performance. The results suggest that the model is robust to changes in m, as only marginal performance differences are observed when m is varied from 2 to 8. (2) **Impact of the shift parameter** n: In contrast, the shift parameter n in the *Top-k shifted by N* mechanism plays a more critical role. Our experiments show that smaller values of n yield better performance, while larger n values lead to a notable drop in both EM and F1 scores. We hypothesize that as n increases, the task of mining hard negatives becomes more challenging, which in turn hampers the fine-tuning process of RADIO.



Example 1: Query: which state is the richest state in nigeria
Ground Truth: ['Lagos State']

Document Reranked by RADIO

1. Niger State is a state in central Nigeria, and it is the largest state in the country in terms of land area. 2. Introduction to Niger State. 3. Lagos State is the most economically important state in the country, containing the city of Lagos, which is Nigeria's largest city. Lagos State is a significant financial center, and if it were a country, it would be the fifth-largest economy in Africa. 4. Ugbo Kingdom is a town in the Ilaje Local Government Area of Ondo State in the southwestern part of Nigeria. Ugbo Kingdom is ranked as the richest king in Nigeria. 5. Rivers State, also known simply as Rivers, is one of the 36 states of Nigeria.

Response Given by RADIO: ✓ Lagos State

Document Reranked by DPA-RAG

1. Niger State is a state in central Nigeria, and it is the largest state in the country in terms of land area. 2. Ugbo Kingdom is a town in the Ilaje Local Government Area of Ondo State in the southwestern part of Nigeria. Ugbo Kingdom is ranked as the richest king in Nigeria. 3. Introduction to Niger State. 4. Kano is located in the northern part of Nigeria and is an important commercial and economic center of Nigeria. 5. The official language of Nigeria is English. The religious composition of Nigeria is roughly divided into Christians in the southern region and Muslims in the northern region.

Response Given by DPA-RAG: ✗ Can not give a answer.

Example 2: Query: who is the highest selling r&b artist of all time
Ground Truth: ['Rihanna', 'Michael Jackson']

Document Reranked by RADIO

1. R. Kelly is recognized as one of the best-selling music artists in the US. 2. Record sales definitions. Michael Jackson's 1982 album "Thriller" is still the best-selling album of all time. 3. R. Kelly is considered one of the most successful R&B artists of the past 25 years. 4. 1990s in Music. Janet Jackson, Michael Jackson, Whitney Houston, and Mariah Carey are among the best-selling music artists, particularly in the 1990s when they brought contemporary R&B to the global stage. 5. 2000s in Music. Destiny's Child, composed of Beyoncé, Michelle Williams, and Kelly Rowland, is the most successful female R&B group in history, selling over 50 million records globally in the 2000s.

Response Given by RADIO: ✓ Michael Jackson

Document Reranked by DPA-RAG

1. R. Kelly is considered one of the most successful R&B artists of the past 25 years. 2. R. Kelly was one of the most successful male R&B artists of the 1990s and one of the best-selling music artists of all time. 3. R. Kelly performed on the same stage as other hip-hop artists such as Nas, Sean Combs, and The Notorious B.I.G. 4. 2000s in Music. Destiny's Child, composed of Beyoncé, Michelle Williams, and Kelly Rowland, is the most successful female R&B group in history, selling over 50 million records globally in the 2000s. 5. Rhythm and Blues.

Response Given by DPA-RAG: ✗ R. Kelly

Figure 3: Case study on NQ dataset.

A.7 Case Study

To intuitively illustrate the effectiveness of RADIO, we select examples from the NQ dataset to compare the documents reranked by RADIO with those reranked by a novel baseline, DPA-RAG, as well as their responses. In Figure 3 Example 1, the query asks, "*Which state is the richest state in Nigeria?*" RADIO successfully ranks information about Lagos State's economic and financial status, which relates to the correct answer, among the top-3 documents. In contrast, DPA-RAG fails to identify documents relevant to answering the query, and cannot provide a valid response. In Figure 3 Example 2, the query is, "*Who is the highest-selling R&B artist of all time?*" RADIO prioritizes documents containing information about the correct answer, Michael Jackson, and effectively highlights key terms such as "R&B" and "best-selling." However, DPA-RAG misinterprets the query's constraints, retrieving documents that either overlook the R&B artist specification or fail to consider the time span, resulting in an incorrect response. These examples demonstrate that RADIO enhances RAG by providing a more efficient and accurate reranking. It selects contextually appropriate documents, enabling the generator to infer correct answers.

A.8 More Results on MMLU

In this section, we give the complete experimental results on MMLU. Specifically, we use the NQ dataset and TriviaQA dataset as source dataset to fine-tune rerankers and evaluate them in MMLU. The results are shown in Table 11.

A.9 More Results on Hyperparameter Analysis

Figure 4 illustrates the trend of RADIO's performance across various MMLU categories as the integration coefficient α increases. For Humanities, Social Sciences, and Other, the trends are consistent: as α increases, the performance metrics first improve and then decline. However, the STEM category shows a unique pattern, with metrics initially decreasing as α grows, followed by an improvement. This divergence may be attributed to the fine-tuning dataset (NQ), which shares a closer distribution with Humanities, Social Sciences, and Other categories, while differing significantly from the STEM category.

Table 11: Experimental results on MMLU. EM is reported as the metric. The source datasets used to fine-tune rerankers are Open-domain QA datasets NQ and TriviaQA.

Method	MMLU (Source Dataset NQ)					MMLU (Source Dataset TriviaQA)				
	Humanities	Social	STEM	Other	ALL	Humanities	Social	STEM	Other	ALL
Base	0.4089	0.6867	**0.5147**	0.6650	0.5502	0.4089	0.6867	**0.5147**	0.6650	0.5502
Atlas	0.3985	0.6935	0.5074	0.6563	0.5447	0.3966	0.6822	0.4868	0.654	0.5364
REPLUG	0.4102	0.6854	0.5065	0.6590	0.5473	0.3977	0.6744	0.4821	0.6466	0.5323
RRR	0.4079	0.6913	0.5116	0.6572	0.5484	0.4132	0.6926	0.5005	0.6501	0.5464
ARL2	0.4147	0.7016	0.5106	0.6630	0.5540	0.4012	**0.6951**	0.5011	0.6639	0.5462
DPA-RAG	0.4157	0.701	0.5078	0.6652	0.5541	0.4189	0.6932	0.5062	**0.6746**	0.5552
RADIO (Ours)	**0.4172**	**0.7013**	0.5080	**0.6717**	**0.5562**	**0.4230**	0.6942	0.5090	0.6678	**0.5559**

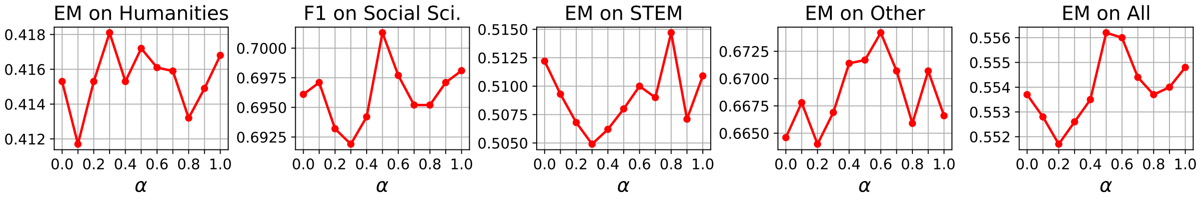


Figure 4: More results of hyperparameter analysis on MMLU.