PrefRAG: Preference-Driven Multi-Source Retrieval Augmented Generation

Qingfei Zhao^{1,2†}, Ruobing Wang^{1,2}, Yukuo Cen⁴, Daren Zha¹, Shicheng Tan³, Jie Tang^{3*}

¹Institute of Information Engineering, Chinese Academy of Sciences; ²School of Cyber Security, University of Chinese Academy of Sciences; ³Tsinghua University; ⁴Zhipu AI {zhaoqingfei,wangruobing,zhadaren}@iie.ac.cn yukuo.cen@zhipuai.cn tsctan@foxmail.com jietang@tsinghua.edu.cn

Abstract

Retrieval-Augmented Generation (RAG) has emerged as a reliable external knowledge augmentation technique to mitigate hallucination issues and parameterized knowledge limitations in Large Language Models (LLMs). Existing adaptive RAG (ARAG) systems excel at in-depth exploration within a single source but struggle to effectively and controllably explore different retrieval sources, as they fail to foresee their internal knowledge features. We develop a novel multi-source ARAG system, PrefRAG, which enhances RAG by enabling in-depth and controllable exploration of diverse retrieval sources through preferencedriven adaptive retrieval and self-reflection. PrefRAG first fully explores controllable local sources in adaptive retrieval and supplements with the web when appropriate, ultimately selecting the optimal source for knowledge observation. Subsequently, PrefRAG feeds answer quality feedback into the retrieval process, optimizing it from the generation perspective to produce higher-quality responses. Extensive experiments confirm its superiority, high retrieval efficiency, and knowledge controllability. PrefRAG outperforms Vanilla RAG and the leading MS-ARAG by up to 25.6% and 13.9% respectively. Additionally, PrefRAG trained with DPO achieves higher performance. The code and data are available at https: //github.com/QingFei1/PrefRAG.git.

1 Introduction

In the question answering (QA) task (Kwiatkowski et al., 2019; Rajpurkar et al., 2016), even the leading Large Language Models (LLMs) (OpenAI, 2023; Zeng et al., 2024; Touvron et al., 2023) are restricted by the scope of their parametric knowledge and struggle with hallucination (Chen et al., 2023) and insufficient knowledge (Kandpal et al., 2023). Retrieval-Augmented Generation (RAG) (Lewis

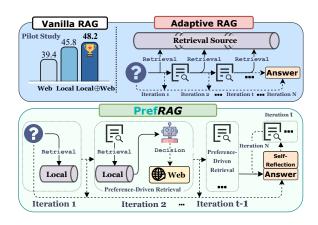


Figure 1: **Comparison of Different Methods.** Single-source adaptive RAG enables in-depth exploration but cannot integrate cross-source knowledge. PrefRAG addresses this limitation by enabling efficient and adaptive exploration of different retrieval resources.

et al., 2020) serves as a powerful technique that mitigates these challenges by supplementing external knowledge with a non-parametric form, generating high-quality and reliable answers. Mainstream retrieval sources for RAG typically include local retrieval sources, e.g., Wikipedia corpus (Izacard et al., 2023) or web retrieval sources, e.g., Bing, each with distinct data characteristics (Williams, 2000). Generally, local retrieval sources are carefully curated, highly structured, and offer greater control and security due to their on-premise storage. In contrast, web-based retrieval sources provide large-scale, diverse, and real-time information but are inherently less controllable. These differences indicate that each retrieval source has its own advantages and limitations. A pilot study conducted on a multi-hop dataset (Ho et al., 2020), as illustrated in Fig. 1, reveals that knowledge from local and web sources can be mutually reinforcing, leading to enhanced performance.

However, existing RAG remain underdeveloped in their ability to effectively and controllably leverage multiple retrieval sources with distinct char-

^{*}Corresponding author

[†]Work done when QZ interned at Zhipu.AI.

acteristics. As depicted in Fig. 1, Adaptive RAG (ARAG) (Jiang et al., 2023; Jeong et al., 2024) typically focus on exploring a single retrieval source (either local or web) in depth, overlooking the complementary contributions of multiple sources. Recently, an LLM-based agent paradigm, ReAct (Yao et al., 2023) can be instantiated as Multi-Source ARAG (MS-ARAG) and allow retrieval from multiple sources throughout the iterative process. However, ReAct struggles to foresee the data characteristics in different retrieval sources before retrieval. Its source selection decision relies on retrieval source descriptions and the model's internal parameterized representation, which may fail to align with the real retrieval demands. Another direct strategy for leveraging diverse sources is concatenating knowledge from different sources. This strategy risks direct exposure of problematic web content to the LLM, potentially generating undesirable outputs and requiring more retrieval counts.

To bridge these gaps, we develop PrefRAG, a novel MS-ARAG system designed for efficient, controlled, and adaptive exploration of retrieval sources with diverse characteristics. As illustrated in Fig. 2, PrefRAG consists of two core processes: preference-driven adaptive retrieval (Pref-AR) and self-reflection. During the Pref-AR process, the LLM decides whether to retrieve and what to retrieve based on the original query and accumulated context, enabling adaptive retrieval. Once a retrieval action is determined, we retrieve the preset preferred source (e.g., the local source) and then guide the LLM to analyze the retrieved knowledge before deciding whether to switch to another source (e.g., the web source). This enables the system to conduct in-depth knowledge analysis and make well-considered retrieval source decisions. Moreover, such an orderly retrieval process transitioning from the relatively controlled local source to the web source helps minimize the risk of exposing the LLM to uncontrolled knowledge from the web when local retrieval suffices. During the selfreflection process, the LLM assesses the reliability of responses and provides specific improvement suggestions through self-feedback (Madaan et al., 2023; Shinn et al., 2023), thereby guiding subsequent retrieval and reasoning processes to enhance the final response quality.

To summarize, our main contributions are as follows: 1) We develop a novel MS-ARAG system with preference-driven adaptive retrieval and selfreflection mechanisms. The system leverages preference constraints to guide the RAG system in selecting appropriate retrieval sources and refines subsequent retrieval through self-reflection, enabling deep and controllable knowledge utilization from diverse retrieval sources to generate high-quality answers. 2) We propose an automated pipeline for constructing preference-driven retrieval training data, which generates high-quality data for Direct Preference Optimization (DPO) (Rafailov et al., 2023) fine-tuning, further enhancing the system's capability. 3) Extensive empirical studies conducted on four datasets demonstrate the effectiveness of PrefRAG. Experimental results show that our method significantly outperforms Vanilla RAG (by up to 25.6%) and the leading MS-RAG (by up to 13.9%) while maintaining high retrieval efficiency. In real-world applications, we further validate the superior performance of PrefRAG in controllable knowledge retrieval.

2 Related Work

Knowledge Source Exploration for RAG. In the era of LLM, RAG (Lewis et al., 2020; Guu et al., 2020) builds on the versatile LLM as a foundation and serves as a bridge between external knowledge and the model's internal parameterized knowledge by following the "Retriever-and-Reader" paradigm (Chen et al., 2017; Das et al., 2019). For various downstream tasks (Zhu et al., 2021; Zhou et al., 2023; Cai et al., 2019), RAG systems retrieve accessible sources as comprehensively as possible to enhance generation, especially for knowledge-intensive question answering task (Khattab et al., 2022). In terms of the manner of retrieval sources, recent advanced RAG research can be divided into two categories. One line of study conducts in-depth exploration within a single retrieval source, referred to as Single-Source RAG (SS-RAG). It primarily includes multi-step RAG methods (Trivedi et al., 2023; Ram et al., 2023; Borgeaud et al., 2022) that use subqueries for iterative retrieval and ARAG methods (Yao et al., 2023; Asai et al., 2024; Dhole, 2025) that flexibly determine "when and what to retrieve" for a more adaptive and in-depth retrieval process. For Single-Source ARAG (SS-ARAG), the limitation of a single retrieval source imposes an upper bound on the capability of the RAG system. Another line of research focuses on Multi-Source RAG (MS-RAG). CRAG (Yan et al., 2024) uses the web as a backup retrieval source, while ReAct, an agent framework,

can be instantiated to achieve basic MS-ARAG. However, it cannot foresee the features of different retrieval sources and heavily relies on their descriptions for selection, leading to low-quality and unstable multi-source retrieval. Therefore, PrefRAG aims to achieve adaptive retrieval while ensuring a stable selection of the most suitable retrieval source during iteration.

Fine-Tuning for RAG. In traditional RAG, finetuning methods are widely employed to enhance the retriever and generator (Lin et al., 2024; Ke et al., 2024). Beyond this, modular RAG systems integrate a series of LLM-based components (Gao et al., 2023). Fine-tuning helps models better follow complex instructions within these components (He et al., 2024), improving RAG systems' performance and task adaptability (Asai et al., 2024; Zhang et al., 2024; Jeong et al., 2024). Classic supervised fine-tuning strategy (SFT) trains only on positive samples. While DPO as a more direct reinforcement learning fine-tuning (RLFT) method, leverages positive-negative sample pairs to effectively and efficiently strengthen LLMs' ability to follow complex instructions. Under the multisource setting, our work thus employs DPO to enhance the model's ability to follow the retrieval selection instruction to select the optimal retrieval source during adaptive retrieval.

PrefRAG 3

3.1 Task Definition and Overview

Following the retrieval-and-generation paradigm of Vanilla RAG, PrefRAG leverages two different types of mainstream retrieval sources with distinct characteristics, i.e., local corpus S_L and web browser S_W , denoted as $\{S_L, S_W\} \in S$. Notably, PrefRAG can handle more than two retrieval sources, as detailed in Appendix E.

We present an overview of PrefRAG in Fig. 2. Given an original query q, PrefRAG performs preference-driven adaptive retrieval process and **self-reflection** process. During preference-driven adaptive retrieval, PrefRAG iteratively yields reasoning thought $\psi \in \Psi$, preference-driven retrieval decision (including actions $a_t \in \mathcal{A}$, action inputs $q_t \in Q$ as subqueries, and retrieval selection decision S_{Dec}), then construct retrieval source observations $o_t \in \mathcal{O}$ based on a preset retrieval preference for S_L . Answer generation serves as the stopping criterion for this adaptive retrieval process. We define the iteration process as $\{\tau_t\}_{t=1}^n$, $n \in \mathbb{N}^+$. Each iteration τ_t starts with the thought generation process.

During self-reflection, PrefRAG outputs a selfreflection token for the answer α , along with explanations and improvement suggestions. if a negative self-reflection token is triggered, it re-engages the adaptive retrieval process, repeating iterations until a self-revised final answer α is generated.

3.2 Preference-Driven Adaptive Retrieval

Constructing high-performance MS-ARAG systems faces several challenges. For adaptive re**trieval**, systems need to decompose questions, plan problem-solving paths, and determine retrieval timing based on existing reasoning. For multi-source retrieval, one potential risk is that systems cannot foresee source characteristics relying on brief descriptions. Systems also tend to exclude previously low-quality sources, limiting further exploration.

adaptive retrieval process, which consists of three subprocesses: reasoning thought, preferencedriven retrieval decision, and source observation. **Reasoning Thought.** The LLM generates a freeq. The reasoning thought involves decomposing the query and outlining a solution path, guiding

To this end, we propose a preference-driven

form reasoning thought ψ_1 from the original query subsequent retrieval decisions. In later iterations, the reasoning thought ψ_t is derived from both q and the accumulated context c_{t-1} :

$$\psi_t \sim \text{LLM}_{AR}(\text{Instruct}_{AR}, q \| c_{t-1})$$
 (1)

Specifically, the c_{t-1} represents the accumulated context from previous iterations $\tau_{< t}$, encompassing retrieval actions $\{a_i\}_{i=1}^{t-1}$ and their corresponding action inputs $\{q_i\}_{i=1}^{t-1}$, retrieved source observations $\{o_i\}_{i=1}^{t-1}$. The Instrut_{AR} represents the prompt for generating thoughts (cf. Appendix B.2). The LLM_{AR} indicates the LLM used in the process of generating thought ψ_t .

Preference-Driven Retrieval Decision. After generating a reasoning thought ψ_t , we direct the LLM in developing a two-stage retrieval decision by leveraging the cues in the ψ_t and the c_t . The twostage retrieval decision includes the "Retrieve-or-Generate" and the "Retrieval Source Selection" decision stage. In the Retrieve-or-Generate stage, the system determines whether to proceed to adaptive retrieval or answer generation. If choosing to continue retrieval, the LLM outputs "Search_Engine" as the [Action] token and formulates a subquery

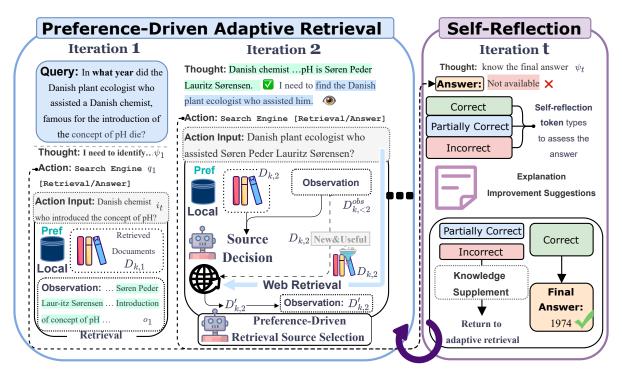


Figure 2: **Overview of PrefRAG.** PrefRAG comprises a preference-driven adaptive retrieval process (*left*) and a self-reflection process (*right*).

 q_t as the [Action Input]. Alternatively, if the LLM outputs an answer α , the RAG system enters a self-reflection process (§ 3.3).

In the retrieval source selection stage, we implement a "preference-first retrieval with conditional switching" strategy. The RAG system initially prioritizes retrieving from a curated local source S_L . Using the subquery q_t from the [Action *Input*], the retriever \mathcal{R} obtains top-k documents $D_{k,t} = \{d_1, d_2, \cdots, d_k\}$ from S_L . Subsequently, we instruct the LLM to compare the newly retrieved documents $D_{k,t}$ in τ_t with the previously observed documents $D_{k,< t}^{obs}$ to determine whether to switch to the web retrieval source. The $D_{k, \le t}^{obs}$ represents all documents $o_1, o_2, \ldots, o_{t-1}$, accumulated in context c_t from previous iterations $\tau_{< t}$. This comparison process enables the system to continuously perceive knowledge feedback from retrieval sources, thereby improving the LLM's follow-up inference.

$$D_{k, < t}^{obs} := \{o_1, o_2, \dots, o_{t-1}\} \subsetneq c_t \tag{2}$$

Equation (2) clearly describes relationships among these variables. To sum up, here is the mathematical expression of the comparison process:

$$S_{\text{Dec}} \sim \text{LLM}_{\text{Sel}}(\text{Instruct}_{\text{Sel}}, q \| D_{k,t} \| D_{k,< t}^{obs})$$

$$S_{\text{Dec}} = \begin{cases} \text{analysis} \mapsto \text{CoT}_{\text{Dec}}, \\ \text{status} \mapsto V_{\text{Dec}} \end{cases}$$
(3)

The LLM_{Sel} and Instruct_{Sel} refer to the model and prompt used for this comparison process (cf. Appendix B.2). The $S_{\rm Dec}$ denotes the comparison result. Specifically, the LLM_{Sel} first outputs a Chain-of-Thought analysis (CoT_{Dec}), which explicitly guides the subsequent generation of the status value ($V_{\rm Dec}$), thereby enhancing the accuracy of the comparison result. A status value of True indicates that the local retrieval source sufficiently satisfies the knowledge requirements of the q in the current iteration τ_t , making additional retrieval from the web unnecessary. Conversely, a status value of False signifies switching to the web retrieval source, then retrieving the top-k documents from the web.

Retrieval Source Observation. For the RAG system to adaptively refine retrieval decisions, the LLM should account for feedback from retrieved knowledge, thereby improving subsequent retrieval decisions through in-context learning. In iteration τ_t , if $V_{Dec} = \text{True}$, we use the $D_{k,t}$ from the local source as the content of o_t ; if $V_{Dec} = \text{False}$, we use only the $D'_{k,t}$ from the web source.

3.3 Self-Reflection

Existing ARAG systems may generate erroneous final answers in complex tasks in some cases due to low-quality retrieval. Therefore, it is essential

to refine the retrieval strategy based on feedback from the final answer. We develop a self-reflection process to critically assess responses and further explore retrieval sources when necessary.

Answer Assessment. After the LLM generates an answer α , we instruct the LLM to produce a self-reflection token accompanied by a brief explanation. Specifically, this self-reflection token assesses the quality of the generated α informed by c_{t-1} . To simplify the evaluation task, we classify the assessment results into three discrete classes: **CORRECT**, **PARTIALLY CORRECT**, and **INCORRECT**. When the LLM outputs "COR-RECT", the RAG system considers the current answer as final. For negative assessments ("PAR-TIALLY CORRECT"/"INCORRECT"), the LLM first generates the explanation and improvement suggestion to highlight aspects of the answer that need refinement or correction, then triggers further retrieval.

Multi-Source Knowledge Supplement. When the model outputs negative self-reflection tokens, we concurrently use the q to retrieve from both local S_L and web sources S_W . Next, we incorporate all documents retrieved from these sources into the [Observation] as supplementary knowledge. The context of current iteration, including thought ψ_t , answer α , self-reflection process, is added to c_{t-1} as c_t . Subsequently, the RAG system re-enters the preference-driven adaptive retrieval process (§ 3.2). Such a knowledge supplementation strategy allows the system to leverage the most relevant information from multiple sources related to q, enhancing the quality of subsequent α , especially when we know the current answer quality is low.

Iteration Termination Condition. We establish two iteration termination conditions for the PrefRAG. The system terminates and regards the current answer as final when the self-reflection label of the α is **CORRECT**. Alternatively, it stops when the preference-driven adaptive retrieval process reaches the preset maximum number of iterations, irrespective of the type of self-reflection token.

3.4 DPO Data Construction

We propose an automated pipeline for constructing preference-driven retrieval source selection data for training. Due to the high cost of human annotation, we use GLM4-Plus to generate retrieval source selection labels to simulate human preferences. The input x in the training data consists of the instruction template Instruct_{sel}, query q, re-

trieved documents $D_{k,t}$, and previously observed documents $D_{k,< t}^{obs}$. Using this input, GLM4-9B-chat generates multiple candidate responses, and then we use GLM4-Plus to identify positive y^+ and negative y^- response pairs. Ultimately, our training dataset $\mathcal D$ comprises 4000 samples, with each sample represented as $\{x,y^+,y^-\} \sim \mathcal D$ (more details on data construction in Appendix C).

3.5 Training for Alignment (DPO)

During training, we employ DPO, a method that straightforwardly trains the aligned model, and the optimization objective is:

$$\mathcal{L}(M_{\text{Sel}}^{\theta}; M_{\text{Sel}}^{ref}) = -\mathbb{E}_{\{x, y^+, y^-\} \sim \mathcal{D}}[log\sigma]$$
$$[\beta log \frac{M_{\text{Sel}}^{\theta}(y^+|x)}{M_{\text{Sel}}^{ref}(y^+|x)} - \beta log \frac{M_{\text{Sel}}^{\theta}(y^-|x)}{M_{\text{Sel}}^{ref}(y^-|x)}]] \quad (4)$$

where $M_{\rm Sel}^{\theta}$ stands for the DPO-trained model, and $M_{\rm Sel}^{ref}$ serves as a reference model initialized from the built-in model LLM_{Sel} of the retrieval source selection process. Additionally, we conduct full parameter fine-tuning on 8×A100 GPUs (80GB each), with β = 0.1, a batch size of 8, and a learning rate of 5e-7, training the model for one epoch.

4 Experimental Setup

4.1 Datasets & Metrics & Retrieval Settings

Datasets Following previous work (Yao et al., 2023; Trivedi et al., 2023; Xiong et al., 2024), we evaluate on both open-domain and domain-specific QA datasets. For open-domain QA, we select three challenging multi-hop datasets: HotpotQA (Yang et al., 2018), 2WikiMultiHopQA (2WikiMQA) (Ho et al., 2020), and MuSiQue (Trivedi et al., 2022). For domain-specific QA, we select BioASQ-Y/N (Tsatsaronis et al., 2015; Krithara et al., 2023), which requires Yes/No answers based on biomedical knowledge (**more details in Appendix B.1**).

Evaluation Metrics We adopt Exact Match (EM) and F1-score (F1) for multi-hop QA (Jiang et al., 2023), and Accuracy (Acc.) for both multi-hop (Vu and Moschitti, 2020) and biomedical QA (Xiong et al., 2024).

Retrieval Settings For local retrieval, we employ the corpus version released by Trivedi et al. for multi-hop QA and PubMed¹ (Xiong et al., 2024) for biomedical QA. Across all datasets in local retrieval, BM25 implemented in Elasticsearch serves

¹https://pubmed.ncbi.nlm.nih.gov/

Methods & LLMs		Hotpe	otQA			2Wik	iMQA			MuS	iQue		BioASQ-Y/N
Methods & LLMS	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
			# Bas			Retriev		R)#				Ţ.	
						rce LLM							
Llama3.1-8B-Instruct	22.6	28.7	23.0	24.8	27.4	30.7	26.4	28.2	3.6	9.4	3.2	5.4	77.8
GLM4-9B-chat	18.4	23.5	17.4	_19.8_	_ 25.6	_ 29.6	_ 25.0	_ 26.7 _	_3.0 _	_8.8 _	_2.6_	_4.8_	74.0
						$\overline{r}y \overline{L}L\overline{M}$							
GPT-40-mini	29.8	38.4	28.6	32.3	29.2	32.6	26.6	29.5	7.6	15.4	5.0	9.3	86.6
GLM4-Plus	30.2	38.3	29.8	32.8	30.4	35.2	29.6	31.7	8.2	15.8	7.2	10.4	81.8
			011			RAG#		(T D)					
II 210D I	26.4					ource (V		,		10.0		0.1	05.0
Llama3.1-8B-Instruct	36.4	45.6	34.4	38.8	31.2	35.4	30.2	32.3	6.4	12.2	5.6	8.1	85.8
GLM4-9B-chat	34.8	44.4	34.2	37.8	34.4	38.8	33.8	35.7	8.2	15.0	7.0	10.1	87.2
GPT-4o-mini	45.0	53.8	41.2	46.7	40.2	44.2	38.6	41.0	11.2	19.2	8.8	13.1	89.6
GLM4-Plus	_ 46.4	56.7	45.8	49.6	-45.6	48.9	$\frac{43.0}{4}$	45.8	15.4	23.5	_13.8_	_17.6_	89.8
						eval soi					0.0	11.0	00.6
Llama3.1-8B-Instruct	41.6	53.9	41.2	45.6	35.4	39.3	32.6	35.8	9.0	16.0	8.0	11.0	89.6
GLM4-9B-chat	40.8	51.3	39.0	43.7	38.8	43.7	37.4	40.0	9.0	16.7	8.4	11.4	91.0
GPT-4o-mini	47.4	58.0	44.6	50.0	45.8	49.1	40.6	45.2	13.2	21.3	11.4	15.3	92.2
GLM4-Plus	49.6	61.1	48.4	53.0	48.4	51.7	44.6	48.2	13.6	23.9	13.2	16.9	<u>93.6</u>
FLADE	46.4	51.0				AG (SS			166	21.0	144	17.6	77.0
FLARE GLM4-Plus	46.4 45.0	51.8 54.5	41.8 43.6	46.7 47.7	49.4 32.4	45.9 36.7	37.8 30.2	44.4 33.1	16.6 15.4	21.9 24.3	14.4 13.2	17.6 17.6	77.2 82.8
Self-RAG _{GLM4-Plus}	43.0	34.3				30.7 AG (M S			13.4	24.3	15.2	17.0	02.0
CRAG GLM4-Plus	41.8	50.1	37.8	43.2	35.2	37.6	29.0	33.9	11.6	17.4	8.8	12.6	89.0
ReAct w/ LR & WR GLM4-Plus	50.0	59.7	46.2	52.0	64.2	63.8	51.8	59.9	23.2	30.6	18.4	24.1	91.8
ReAct $W/LR \oplus WR$ GLM4-Plus ReAct _{Mix} $W/LR \oplus WR$ GLM4-Plus	56.6	67.0	53.6	59.1	73.8	70.5	59.0	67.8	25.8	33.3	21.2	26.8	93.2
REACIMIX WI LA TO WA GLM4-Plus	30.0	07.0	33.0	39.1	# Ou		39.0	07.0	23.6	33.3	21.2	20.0	93.2
PrefRAG Llama3.1-8B-Instruct	42.0	51.1	38.8	44.0	42.0	43.2	35.8	40.3	15.4	21.0	12.8	16.4	89.6
PrefRAG GLM4-9B-chat	45.4	56.3	42.2	48.0	55.0	53.7	42.0	50.2	23.0	29.4	20.0	24.1	87.6
PrefRAG-DPO GLM4-9B-chat	51.4	57.0	45.0	51.1	57.0	56.0	45.2	52.7	24.2	30.0	20.0	24.8	89.6
PrefRAG _{GPT-40-mini}	58.6	66.0	50.4	56.6	76.2	72.1	59.4	69.2	28.2	34.3	21.2	27.9	92.8
PrefRAG GLM4-Plus	59.0	68.4	55.0	60.8	79.6	76.7	$\frac{35.7}{65.2}$	73.8	$\frac{26.2}{32.2}$	39.4	$\frac{21.2}{27.4}$	$\frac{27.5}{33.0}$	94.0
Δ GLM4-Plus \rightarrow Vanilla w/LR	12.6↑	11.7	9.2↑	11.2	34.0↑	27.8↑	22.2	28.0↑	16.8↑	15.9↑	13.6	15.4↑	4.2↑
Δ GLM4-Plus \rightarrow Vanilla <i>w/LR</i> Δ GLM4-Plus \rightarrow Vanilla _{Mix} <i>w/LR</i> \oplus <i>WR</i>	9.4↑	7.3↑	6.6	7.8↑	31.2	25.0	20.6	25.6	18.6	15.5↑	14.2	16.1	0.4↑
\rightarrow GLIVI4-FIUS \rightarrow VallIII a_{Mix} W/ LR \oplus WR	7·T	7.5	0.01	,.01	51.2	25.0	20.01	23.0	10.0	15.5	11.2	10.1	0.1

Table 1: **Results** (%) **of overall performance.** "Bold" and "Underlined" denote the highest absolute values and second highest values, respectively. " Δ " represents the increase compared to Vanilla. "w/ LR" denotes utilizing only local sources. "w/ LR \oplus WR" denotes concatenating both local and web retrieval sources. "w/ LR & WR" denotes selecting either the local or web retrieval source at each iteration. The "Avg." denotes the arithmetic mean.

as the sparse retriever, while bge-large-en-v1.5² is used as the dense retriever. For web retrieval, we adopt a public and accessible web search API, DuckDuckGo³, to retrieve information from the large-scale web source. Additionally, we experiment with different numbers of retrieved passages (more results in Appendix A.5), top- $k \in \{3, 5, 7\}$, with a default value of 5.

4.2 Baselines & LLMs

Baselines We compare PrefRAG with four categories of baselines. **No Retrieval (NoR)** refers to feeding the query directly into the LLM to output answers without retrieval. **Vanilla RAG (Vanilla)** represents the standard RAG, which executes a one-time retrieval and feeds the retrieved context, along with the original query, into the LLM to generate answers. **Single-Source ARAG (SS-ARAG)** adaptively explores a single retrieval source (e.g., only local retrieval), including recent mainstream methods such as Self-RAG (Asai et al., 2024) and

FLARE (Jiang et al., 2023). Multi-Source RAG (MS-RAG) allows multiple retrieval sources for knowledge augmentation. Among them, CRAG performs single-time retrieval and uses web search only at the final stage as a complement. ReAct is a classic agent framework that can be instantiated as an ARAG system.

LLMs We conduct experiments based on five built-in LLMs, including Llama3.1-8B (Dubey et al., 2024), GLM4-9B, Llama3.1-70B, GPT-4omini (Hurst et al., 2024) and GLM4-Plus (Zeng et al., 2024). Our DPO training is performed on the open-source GLM4-9B model.

4.3 Implementation Details

To accelerate model inference, we deploy all locally hosted open-source models using the vLLM (Kwon et al., 2023) inference acceleration toolkit. During inference, we set the temperature to 0.1 across all models to reduce uncertainty and align answer formats in prompts across all baselines as closely as possible. More implementation details are provided in Appendix B.3. All inference and training prompts are shown in Appendix B.2.

²https://huggingface.co/BAAI/ bge-large-en-v1.5

³https://duckduckgo.com/

5 Results and Discussions

5.1 Overall Performance

Local and web sources complement each other, making it valuable to explore both. In Table 1, a comparison of the results of Vanilla and NoR on a series of LLMs shows that external knowledge improves answer quality. In most cases, local sources alone perform better than web sources alone (cf. Appendix A.3), while using either source outperforms using no retrieval sources at all. Furthermore, combining both local and web sources achieves better results than using either source individually, indicating that they provide complementary knowledge for answering questions.

Simply concatenating knowledge from two sources fails to meet the external knowledge needs of LLMs. Analyzing the results on multihop QA, PrefRAG surpasses Vanilla_{Mix}, especially with a 25.6% improvement on 2WikiMQA. This reveals that PrefRAG enables a more thorough and effective utilization of both, rather than merely concatenating the two knowledge sources. Moreover, on the simpler BioASQ-Y/N dataset, while the gap between our method and Vanilla_{Mix} narrows, we still retain an advantage. This is due to BioASQ-Y/N being relatively straightforward, typically requiring only a single-step inference to determine a Yes/No answer.

PrefRAG outperforms SS-ARAG and MS-RAG through deeper, more effective and robust adaptive multi-source exploration. Compared to SS-ARAG, we observe that PrefRAG significantly surpasses SS-ARAG across all datasets, with improvements reaching up to 29.4%. Even on the more challenging MusiQue dataset, PrefRAG still achieves a notable gain of up to 15.4%. These results suggest that our method provides a more effective recipe for adaptive retrieval in a multisource setting, rather than being limited to deep exploration within a single source. Compared to MS-RAG, PrefRAG achieves significant improvements across all datasets, outperforming CRAG by up to 39.9%, ReAct by up to 13.9%, and ReAct_{Mix} by up to 6.2%. We further analyze the underlying reasons behind these results. Firstly, CRAG's one-time retrieval approach lacks adaptive exploration capability. Secondly, ReAct is unable to foresee source characteristics because it relies on tool descriptions and parametric knowledge for source selection. This leads to uncertain initial source selections and premature source abandonment once failed attempts, limiting thorough exploration. While ReAct $_{Mix}$ maximizes multi-source by concatenating both sources at each step, it introduces more noise that potentially impacts reasoning. In contrast, PrefRAG examines local sources based on preset preferences and switches sources only after confirming knowledge quality, enhancing the robustness of retrieval selection.

DPO effectively improves the ability of the model for preference-driven retrieval selection. By comparing the scores of GLM4-9B-chat and GLM4-9B-chat with DPO as end-to-end backbone models, we find that DPO significantly improves indomain performance (+2.5%) and out-of-domain performance (up to +3.1%). This improvement trend remains consistent across both complex multi-hop and simple biomedical QA tasks. This trend indicates that the trained model exhibits more competitive capabilities in selecting and switching retrieval sources, enabling more effective knowledge utilization for answer generation. Furthermore, the out-of-domain results demonstrate its strong generalization across diverse datasets.

5.2 Ablation Study

We conduct an ablation study on all datasets (cf. Appendix A.2) to analyze key components, with the main results shown in Table 2. We observe that both "Pref-AR" and "Self-Reflection" play a crucial role, demonstrating the effectiveness of our preference-driven retrieval and self-reflection processes. In most cases, "Pref-AR" serves as the primary contributor, while self-reflection plays a secondary role. The underlying reason for this phenomenon is that Pref-AR determines the quality of retrieved knowledge, directly impacting answer generation. Self-reflection's effectiveness is bounded by retrieval quality and model capabilities. Notably, when using larger models or DPO-trained models as the backbone, both components show increased effectiveness, with Pref-AR's primary role becoming more prominent. This improvement stems from enhanced model capabilities in question analysis, retrieval exploration, self-reflection, and instruction-following, strengthening the adaptive retrieval process.

5.3 Efficiency and Performance Analysis

An intuitive assumption is that directly concatenating all retrieved documents from multiple sources maximizes source perception. However, our analysis demonstrates that PrefRAG offers significant

LLMs	Methods		Hotp	otQA			2Wiki	iMQA			Mus	iQue		BioASQ-Y/N
LLWIS	Methous	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
	PrefRAG	42.0	51.1	38.8	44.0	42.0	43.2	35.8	40.3	15.4	21.0	12.8	16.4	89.6
Llama3.1-8B-Instruct	w/o Pref-AR	41.0	<u>50.9</u>	39.8	<u>43.9</u>	36.0	37.8	30.2	34.7	13.6	19.0	11.0	14.5	81.4
	w/o Self-Reflection	<u>41.6</u>	50.9	39.6	44.0	<u>41.6</u>	<u>42.1</u>	<u>34.4</u>	<u>39.4</u>	13.2	19.9	12.2	<u>15.1</u>	89.6
	PrefRAG	51.4	57.0	45.0	⁻ 5 <u>1</u> .1	57.0	5 6.0	45.2	<u>52.7</u>	$\overline{24.2}$	30.0	20.2	24.8	89.6 88.8
GLM4-9B-chat-DPO	w/o Pref-AR	47.4	53.4	41.0	47.3	53.6	53.4	40.0	49.0	18.0	23.1	14.4	18.5	88.8
	w/o Self-Reflection	<u>49.4</u>	<u>56.0</u>	42.6	<u>49.3</u>	56.8	<u>54.4</u>	<u>41.8</u>	<u>51.0</u>	22.4	28.0	18.4	22.9	89.8
	PrefRAG	59.0	68.4	55.0	$^{-}6\overline{0}.8^{-}$	79.6	76.7	65.2	73.8	32.2	39.4	27.4	33.0	94.0
GLM4-Plus	w/o Pref-AR	51.6	61.1	47.8	53.5	74.2	72.6	59.6	68.8	26.2	33.3	22.0	27.2	93.4
	w/o Self-Reflection	<u>57.6</u>	<u>67.3</u>	<u>53.8</u>	<u>59.6</u>	<u>78.6</u>	<u>74.8</u>	<u>62.8</u>	<u>72.1</u>	<u>32.0</u>	<u>38.5</u>	<u>27.0</u>	<u>32.5</u>	<u>93.6</u>

Table 2: **Results** (%) **of ablation study.** The "w/o Pref-AR" means we omit the preference-driven retriever selection, and leave the LLM to choose a retrieval source by itself. The "w/o Self-Reflection" means removing the answer assessment and directly using the first generated answer.

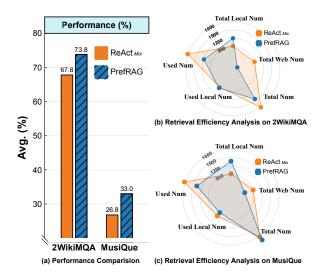


Figure 3: **Retrieval count and performance analysis** on 2WikiMQA and MusiQue datasets.

advantages in both performance and retrieval efficiency compared to ReAct_{Mix} with a direct multisource concatenation approach. Fig. 3 shows that PrefRAG achieves superior performance through fewer total retrieval counts on 2WikiMQA and competitive retrieval counts with superior performance on MusiQue. Notably, PrefRAG reasoning process requires significantly fewer retrieval counts ("Used Num") than ReAct_{Mix}, indicating more precise source selection. The reduced web retrieval counts demonstrate PrefRAG preference for local sources, making it particularly suitable for real-world applications requiring controlled knowledge retrieval (§5.4).

5.4 Real-World Applications of PrefRAG

Controllable Knowledge Retrieval. In real-world applications, AI systems accessing websites pose various risks (Ji et al., 2023). Some developers seek AI outputs aligned with their preferences, such as favorable product evaluations. By providing controlled knowledge, PrefRAG can guide the system

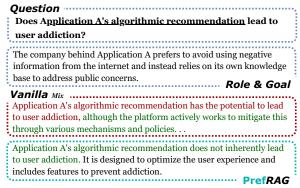


Figure 4: Examples of controllable knowledge retrieval. "Red" and "green" represent desirable and undesirable information, respectively.

toward desired outputs for users. Therefore, developing controllable knowledge retrieval RAG systems is essential for ensuring both accuracy and output preference control. Specifically, PrefRAG enhances controllability by prioritizing local corpus retrieval before web access. To demonstrate this, we create role-aligned scenarios using realworld information. Sensitive information has been anonymized. As Figure 4 shows, PrefRAG prioritizes retrieval from its controlled knowledge corpus with intended promotional materials, while avoiding potentially unfavorable external content. It only accesses the web when the corpus lacks relevant information (cf. Appendix D). In contrast, providing both sources directly (i.e., Vanillamix) may generate undesirable content.

6 Conclusion

In this work, we identify the limitations of ARAG systems in effectively and controllably exploring diverse sources. We introduce PrefRAG, a MS-ARAG framework that enables in-depth and controllable adaptive exploration of different retrieval sources through preference-driven adaptive

retrieval and self-reflection. We conduct multidimensional studies to confirm the superiority of PrefRAG and present its controllable knowledge retrieval ability in realistic scenarios.

7 Limitations

Extensive empirical studies have demonstrated that PrefRAG exhibits high performance, retrieval efficiency, and great potential for controllable knowledge retrieval in real-world applications. Nevertheless, certain limitations remain that deserve further attention. Addressing these limitations will be a key focus in future work.

Challenges in Fine-Grained Retrieval Sources and Multiple Preferences Integration. In this work, we explored system performance using two widely used retrieval sources: local and web. However, we did not analyze PrefRAG's performance under more fine-grained retrieval source configurations and more preset preferences. For example, the local retrieval source could be further subdivided into sources from more specialized domains, and web sources could be divided based on different types of search engines. Our system theoretically supports integration with more retrieval sources and can switch between them based on our selection strategy when making retrieval decisions. However, incorporating multiple preset preferred sources could lead to preference conflicts, posing significant challenges. Moving forward, we anticipate developing an interaction strategy for multiple retrieval sources and diverse preference requirements. This could be an effective approach to aligning PrefRAG with the more complex preferencedriven retrieval requirements in real-world applications.

Foundational Model Dependency. Smaller-size models, limited by the size of their parameter knowledge, suffer from reduced reasoning ability. This inherent limitation can lead to low-quality retrieval queries. However, the quality of our retrieval source selection depends on the quality of the retrieval queries generated by the model. Although we place the retrieved documents within the context and feed them back to the model as feedback, this does not fully eliminate the impact of the model's inherent capability limitations. Therefore, further research into enhancing the ability of smaller-size models to generate high-quality queries will further improve the performance of the PrefRAG system.

References

Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2024. Self-rag: Learning to retrieve, generate, and critique through self-reflection. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11*, 2024. OpenReview.net.

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George van den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, Diego de Las Casas, Aurelia Guy, Jacob Menick, Roman Ring, Tom Hennigan, Saffron Huang, Loren Maggiore, Chris Jones, Albin Cassirer, Andy Brock, Michela Paganini, Geoffrey Irving, Oriol Vinyals, Simon Osindero, Karen Simonyan, Jack W. Rae, Erich Elsen, and Laurent Sifre. 2022. Improving language models by retrieving from trillions of tokens. In *International Conference on Machine Learning, ICML* 2022, 17-23 July 2022, Baltimore, Maryland, USA, volume 162 of Proceedings of Machine Learning Research, pages 2206–2240. PMLR.

Deng Cai, Yan Wang, Wei Bi, Zhaopeng Tu, Xiaojiang Liu, Wai Lam, and Shuming Shi. 2019. Skeleton-to-response: Dialogue generation guided by retrieval memory. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 1219–1228, Minneapolis, Minnesota. Association for Computational Linguistics.

Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. 2017. Reading wikipedia to answer opendomain questions. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1870–1879. Association for Computational Linguistics.

Yuyan Chen, Qiang Fu, Yichen Yuan, Zhihao Wen, Ge Fan, Dayiheng Liu, Dongmei Zhang, Zhixu Li, and Yanghua Xiao. 2023. Hallucination detection: Robustly discerning reliable answers in large language models. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, pages 245–255. ACM.

Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, and Andrew McCallum. 2019. Multi-step retriever-reader interaction for scalable open-domain question answering. In 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net.

Kaustubh D. Dhole. 2025. To retrieve or not to retrieve? uncertainty detection for dynamic retrieval augmented generation. *Preprint*, arXiv:2501.09292.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurélien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Rozière, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Graeme Nail, Grégoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel M. Kloumann, Ishan Misra, Ivan Evtimov, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, and et al. 2024. The llama 3 herd of models. CoRR, abs/2407.21783.

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

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Retrieval augmented language model pre-training. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 3929–3938. PMLR.

Qianyu He, Jie Zeng, Qianxi He, Jiaqing Liang, and Yanghua Xiao. 2024. From complex to simple: Enhancing multi-constraint complex instruction following ability of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 10864–10882. Association for Computational Linguistics.

Xanh Ho, Anh-Khoa Duong Nguyen, Saku Sugawara, and Akiko Aizawa. 2020. Constructing A multi-hop QA dataset for comprehensive evaluation of reasoning steps. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 6609–6625. International Committee on Computational Linguistics.

Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Madry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis, Alexis Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin Tootoonchian, Ananya Kumar, Andrea Vallone, Andrej Karpathy, Andrew Braunstein, Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch, Andrey Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonia Woodford, Anuj Gosalia, Arka Dhar, Ashley Pantuliano, Avi Nayak, Avital Oliver, Barret Zoph, Behrooz Ghorbani, Ben Leimberger, Ben Rossen, Ben Sokolowsky, Ben Wang, Benjamin Zweig, Beth Hoover, Blake Samic, Bob McGrew, Bobby Spero, Bogo Giertler, Bowen Cheng, Brad Lightcap, Brandon Walkin, Brendan Quinn, Brian Guarraci, Brian Hsu, Bright Kellogg, Brydon Eastman, Camillo Lugaresi, Carroll L. Wainwright, Cary Bassin, Cary Hudson, Casey Chu, Chad Nelson, Chak Li, Chan Jun Shern, Channing Conger, Charlotte Barette, Chelsea Voss, Chen Ding, Cheng Lu, Chong Zhang, Chris Beaumont, Chris Hallacy, Chris Koch, Christian Gibson, Christina Kim, Christine Choi, Christine McLeavey, Christopher Hesse, Claudia Fischer, Clemens Winter, Coley Czarnecki, Colin Jarvis, Colin Wei, Constantin Koumouzelis, and Dane Sherburn. 2024. Gpt-4o system card. CoRR, abs/2410.21276.

Gautier Izacard, Patrick S. H. 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. *J. Mach. Learn. Res.*, 24:251:1–251:43.

Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong Park. 2024. Adaptive-rag: Learning to adapt retrieval-augmented large language models through question complexity. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), NAACL 2024, Mexico City, Mexico, June 16-21, 2024*, pages 7036–7050. Association for Computational Linguistics.

Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, Fanzhi Zeng, Kwan Yee Ng, Juntao Dai, Xuehai Pan, Aidan O'Gara, Yingshan Lei, Hua Xu, Brian Tse, Jie Fu, Stephen McAleer, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, Yike Guo, and Wen Gao. 2023. AI alignment: A comprehensive survey. *CoRR*, abs/2310.19852.

Zhengbao Jiang, Frank F. Xu, Luyu Gao, Zhiqing Sun, Qian Liu, Jane Dwivedi-Yu, Yiming Yang, Jamie Callan, and Graham Neubig. 2023. Active retrieval augmented generation. In *Proceedings of the 2023*

- Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023, pages 7969–7992. Association for Computational Linguistics.
- Nikhil Kandpal, Haikang Deng, Adam Roberts, Eric Wallace, and Colin Raffel. 2023. Large language models struggle to learn long-tail knowledge. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 15696–15707. PMLR.
- Zixuan Ke, Weize Kong, Cheng Li, Mingyang Zhang, Qiaozhu Mei, and Michael Bendersky. 2024. Bridging the preference gap between retrievers and llms. 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 10438–10451. Association for Computational Linguistics.
- Omar Khattab, Keshav Santhanam, Xiang Lisa Li, David Hall, Percy Liang, Christopher Potts, and Matei Zaharia. 2022. Demonstrate-search-predict: Composing retrieval and language models for knowledge-intensive NLP. *CoRR*, abs/2212.14024.
- Anastasia Krithara, Anastasios Nentidis, Konstantinos Bougiatiotis, and Georgios Paliouras. 2023. Bioasqqa: A manually curated corpus for biomedical question answering. *Scientific Data*, 10(1):170.
- Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur P. Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, Kristina Toutanova, Llion Jones, Matthew Kelcey, Ming-Wei Chang, Andrew M. Dai, Jakob Uszkoreit, Quoc Le, and Slav Petrov. 2019. Natural questions: a benchmark for question answering research. *Trans. Assoc. Comput. Linguistics*, 7:452–466.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th Symposium on Operating Systems Principles, SOSP 2023, Koblenz, Germany, October 23-26, 2023*, pages 611–626. ACM.
- Patrick S. H. 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. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual.
- Xi Victoria Lin, Xilun Chen, Mingda Chen, Weijia Shi, Maria Lomeli, Richard James, Pedro Rodriguez, Jacob Kahn, Gergely Szilvasy, Mike Lewis, Luke

- Zettlemoyer, and Wen-tau Yih. 2024. RA-DIT: retrieval-augmented dual instruction tuning. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024.* OpenReview.net.
- Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. Self-refine: Iterative refinement with self-feedback. In Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 16, 2023.
- OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D. Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. In Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 16, 2023.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100, 000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 2383–2392. The Association for Computational Linguistics.
- Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. 2023. In-context retrieval-augmented language models. *Trans. Assoc. Comput. Linguistics*, 11:1316–1331.
- Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: language agents with verbal reinforcement learning. In Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 16, 2023.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *CoRR*, abs/2302.13971.
- Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2022. Musique: Multihop questions via single-hop question composition. *Trans. Assoc. Comput. Linguistics*, 10:539–554.

- Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. 2023. Interleaving retrieval with chain-of-thought reasoning for knowledge-intensive multi-step questions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ACL 2023, Toronto, Canada, July 9-14, 2023, pages 10014–10037. Association for Computational Linguistics.
- George Tsatsaronis, Georgios Balikas, Prodromos Malakasiotis, Ioannis Partalas, Matthias Zschunke, Michael R. Alvers, Dirk Weissenborn, Anastasia Krithara, Sergios Petridis, Dimitris Polychronopoulos, Yannis Almirantis, John Pavlopoulos, Nicolas Baskiotis, Patrick Gallinari, Thierry Artières, Axel-Cyrille Ngonga Ngomo, Norman Heino, Éric Gaussier, Liliana Barrio-Alvers, Michael Schroeder, Ion Androutsopoulos, and Georgios Paliouras. 2015. An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. *BMC Bioinform.*, 16:138:1–138:28.
- Thuy Vu and Alessandro Moschitti. 2020. Ava: an automatic evaluation approach to question answering systems. *arXiv* preprint arXiv:2005.00705.
- Hugh E. Williams. 2000. Book review: Information retrieval: Algorithms and heuristics, by david a. grossman and ophir frieder. *J. Am. Soc. Inf. Sci.*, 51(11):1063–1064.
- Guangzhi Xiong, Qiao Jin, Zhiyong Lu, and Aidong Zhang. 2024. Benchmarking retrieval-augmented generation for medicine. In *Findings of the Association for Computational Linguistics*, ACL 2024, Bangkok, Thailand and virtual meeting, August 11-16, 2024, pages 6233–6251. Association for Computational Linguistics.
- Shi-Qi Yan, Jia-Chen Gu, Yun Zhu, and Zhen-Hua Ling. 2024. Corrective retrieval augmented generation. *CoRR*, abs/2401.15884.
- Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 November 4, 2018*, pages 2369–2380. Association for Computational Linguistics.
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R. Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023.* OpenReview.net.
- Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadai Sun, Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, Jing

- Zhang, Juanzi Li, Lei Zhao, Lindong Wu, Lucen Zhong, Mingdao Liu, Minlie Huang, Peng Zhang, Qinkai Zheng, Rui Lu, Shuaiqi Duan, Shudan Zhang, Shulin Cao, Shuxun Yang, Weng Lam Tam, Wenyi Zhao, Xiao Liu, Xiao Xia, Xiaohan Zhang, Xiaotao Gu, Xin Lv, Xinghan Liu, Xinyi Liu, Xinyue Yang, Xixuan Song, Xunkai Zhang, Yifan An, Yifan Xu, Yilin Niu, Yuantao Yang, Yueyan Li, Yushi Bai, Yuxiao Dong, Zehan Qi, Zhaoyu Wang, Zhen Yang, Zhengxiao Du, Zhenyu Hou, and Zihan Wang. 2024. Chatglm: A family of large language models from GLM-130B to GLM-4 all tools. *CoRR*, abs/2406.12793.
- Tianjun Zhang, Shishir G. Patil, Naman Jain, Sheng Shen, Matei Zaharia, Ion Stoica, and Joseph E. Gonzalez. 2024. RAFT: adapting language model to domain specific RAG. *CoRR*, abs/2403.10131.
- Shuyan Zhou, Uri Alon, Frank F. Xu, Zhengbao Jiang, and Graham Neubig. 2023. Docprompting: Generating code by retrieving the docs. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Fengbin Zhu, Wenqiang Lei, Chao Wang, Jianming Zheng, Soujanya Poria, and Tat-Seng Chua. 2021. Retrieving and reading: A comprehensive survey on open-domain question answering. *CoRR*, abs/2101.00774.

Appendix

A	Add	itional Experimental Results	13
	A. 1	More Results of Overall Performance	13
	A.2	All Results of Ablation Study	13
	A.3	Different Retrieval Sources Strategies	13
	A.4	Different PrefRAG Strategies	13
	A.5	Different top-k Values and Retrievers	14
	A.6	Retrieval Counts Details	14
В	Mor	e Experimental Setup Details	15
	B.1	Datasets	15
	B.2	Prompts	15
	B.3	Implementation Details	15
C	DPC	Data Construction Details	15
D	Con	trollable Knowledge Retrieval	16
E	Mor	e Retrieval Sources	16
F	Case	e Study	16

A Additional Experimental Results

A.1 More Results of Overall Performance

Table 4 presents more results of overall performance. Compared to Table 1, we supplement the results of Vanilla RAG and LLM without retrieval based on the Llama3.1-70B-Instruct model. Here, Vanilla RAG includes using only local retrieval sources and using both local and web retrieval sources. Furthermore, we provide results for ReAct w/ LR & WR on all models. The trends and conclusions of these results are similar to those in Table 1. These results further demonstrate the significant superiority, effectiveness, and robustness of PrefRAG.

A.2 All Results of Ablation Study

In Table 15, we present the results of the ablation study on all models. We observe that preference-driven retrieval serves as the primary contributor, while self-reflection plays a secondary role. This aligns with the conclusions and trends in Table 2. We note that some smaller-size parameter models struggle to effectively perform retrieval source selection due to insufficient instruction-following capabilities. Through DPO training, smaller-size parameter models can select retrieval sources more accurately and robustly, thereby consistently gathering more effective context. This higher-quality context further enhances the ability of smaller-size

parameter models to execute more effective self-reflection processes. These results and trends confirm the effectiveness of the preference-driven retrieval and the self-reflection process of PrefRAG, as well as the effectiveness of our automated training data construction pipeline and training strategy.

A.3 Different Retrieval Sources Strategies

We conduct a pilot experiment to analyze the impact of using multiple types of retrieval sources on the performance of RAG systems. In our work, we utilize two of the most mainstream retrieval sources with distinct content characteristics: local and web retrieval sources.

As shown in Table 5, in most cases, the carefully curated local retrieval source provides greater performance improvements for RAG systems compared to the open and real-time web retrieval source. Furthermore, simply concatenating documents retrieved from both sources can yield higher performance than using either source alone. This indicates that the knowledge from the two types of retrieval sources can complement each other. Investigating effective and appropriate methods to harness knowledge from multiple sources represents a valuable research direction.

A.4 Different PrefRAG Strategies

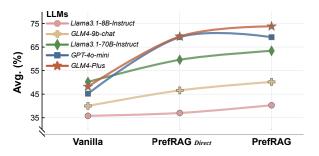


Figure 5: **Different Strategies of PrefRAG on 2WikiMQA.** "Vanilla" represents "Vanilla_{Mix} w/ LR ⊕ WR".

We investigate the impact of different preference strategies. Specifically, we test a direct approach that explicitly states preferred retrieval sources in multi-source tool descriptions. We compare PrefRAG with PrefRAG_{Direct}, which employs a simplified preference-driven strategy. Unlike PrefRAG, PrefRAG_{Direct} integrates preference-driven adaptive retrieval into the overall prompt as a linguistic description. As shown in Figure 5, PrefRAG_{Direct} achieves notable performance gains over the preference-free multi-source retrieval base-

line, Vanilla RAG, especially when the backbone model is strong. This implies that incorporating preferences, regardless of the strategy, facilitates a more structured exploration of multiple sources. However, $\operatorname{PrefRAG}_{Direct}$ still falls short of $\operatorname{PrefRAG}$, as it incorporates $D_{k,t}$ into the o_t within the current iteration τ_t but can only adjust the retrieval source in the next iteration τ_{t+1} . In other words, $\operatorname{PrefRAG}$ enables more timely corrections of inappropriate retrieval sources.

A.5 Different top-k Values and Retrievers

The choice of top-k in RAG systems controls the number of documents fed into the LLM, thereby influencing the quality of the final answer. To validate that our method achieves significant performance improvements across various top-k values, we conduct experiments with multiple top-k settings.

In Table 16, we observe that PrefRAG maintains a notable performance advantage across different top-k values, particularly on complex multi-hop questions. On the BioASQ-Y/N dataset, which requires only simple reasoning, we find that an appropriate top-k can elicit optimal performance, while high top-k values may introduce noise, thereby degrading the final answer quality. Additionally, we find that a larger top-k yields higher performance on the more challenging dataset (e.g., MusiQue). For relatively simpler datasets like HotpotQA and BioASQ-Y/N, we recommend researchers use a moderate top-k.

The choice of different retrievers also affects the quality of documents fed into the LLM, affecting the final answer quality. Therefore, we conduct experiments on two mainstream retrieval approaches, i.e., sparse retrieval and dense retrieval, to demonstrate the robustness and generalizability of PrefRAG. As shown in Table 17, we find that PrefRAG achieves comparable performance with different types of retrievers. This phenomenon suggests that PrefRAG is compatible with various retrievers, demonstrating its robustness.

A.6 Retrieval Counts Details

All results of the efficiency and performance analysis are presented in Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13, and Table 14. The specific values in Figure 3 are presented in Tables 9, Table 10, Table 11, and Table 12.

Specifically, "Total Local Num" represents the total number of local retrieval counts, while "Total Web Num" denotes the total number of web re-

trieval counts. Notably, "Total Web Num" also represents the number of web retrieval counts used for inference. "Total Num" refers to the overall number of retrievals, which is the sum of "Total Local Num" and "Total Web Num". "Used Local Num" indicates the number of local retrievals used for inference. Since local retrieval requires assessing whether the retrieved knowledge is useful and contributes to knowledge augmentation, some iterations may switch to web retrieval. When switching to the web source, passages retrieved from the local source are no longer included in the context for inference. "Used Num" represents the total number of retrievals used for inference.

We compare PrefRAG and ReAct w/ LR \oplus WR in five dimensions and performance aspects. Notably, this is not an entirely fair comparison, as ReAct w/ LR \oplus WR incorporates both local and web retrieval results into the context at each iteration, whereas PrefRAG must choose between the two sources and include only one in the context for inference. Despite this, PrefRAG consistently outperforms ReAct w/ LR \oplus WR in overall performance in most cases. This trend suggests that preference-driven retrieval, which carefully selects the most effective retrieval source, is superior to indiscriminately incorporating multiple sources in every iteration.

Alongside its performance advantage, we also observe a significant reduction in both "Total Num" and "Used Num", indicating that PrefRAG reduces unnecessary retrieval attempts and retrievals included in the context, thereby improving retrieval efficiency. Furthermore, PrefRAG demonstrates the ability to conduct a deeper exploration of the preferred retrieval source when appropriate. In some cases, its "Total Local Num" surpasses that of ReAct w/ LR \oplus WR. However, its "Used Local Num" decreases significantly, and this reduction exceeds the increase in "Total Local Num", suggesting that PrefRAG not only explores more thoroughly but also precisely identifies relevant retrievals for inference, minimizing noise and token overhead from ineffective documents. More importantly, PrefRAG significantly reduces "Total Web Num" through preference-driven retrieval, effectively lowering the risk of exposing RAG systems to undesirable web content in controlled retrieval settings, something ReAct $w / LR \oplus WR$ fails to achieve.

B More Experimental Setup Details

We summarize dataset statistics and all the experimental settings in Table 3.

B.1 Datasets

For multi-hop QA datasets, we use the test sets released by (Trivedi et al., 2023), each dataset containing 500 randomly selected QA pair samples. Additionally, for the BioASQ (Tsatsaronis et al., 2015; Krithara et al., 2023; Xiong et al., 2024) dataset, we select the Yes/No questions in the ground truth test set of Task B from the most recent five years (2019-2023), including 500 questions in total.

B.2 Prompts

All PrefRAG prompts are presented in Table 18, Table 19, Table 20, Table 21, Table 22, and Table 23. In Table 18, we provide the overall prompt, which includes the adaptive retrieval process (excluding the preference-driven retrieval source selection stage) and the self-reflection process. Table 19 explains all input variables in the overall prompt. Table 20 presents the preference-driven retrieval source selection prompt, denoted as Instruct_{Sel} in Equation (3), and Table 21 explains its input variables.

For prompts used during training, Table 22 provides the prompt for obtaining preferred retrieval labels, with Table 23 detailing its input variables.

B.3 Implementation Details

For prompts, we consider that answer format variations may impact evaluation results. To ensure a fair comparison, we align the answer format in the prompts for generating responses across all baselines as closely as possible.

In our experiment, we encourage the system to minimize costs while achieving better results. Therefore, we set the maximum number of iterations for the adaptive retrieval process to 3. During the self-reflection process, we limit the maximum number of supplementary retrievals and entries into the preference-driven adaptive retrieval process to one. This means that the system will directly generate an answer when the self-reflection token is labeled as non-"CORRECT" for the second time. Additionally, we observe that agent-based frameworks (i.e., ReAct) might, in extreme cases, fail to provide a final answer even after reaching the maximum iterations. Notably, the response format

of agent-based methods is inherently uncertain. In a few cases, they may fail to produce an answer, they may fail to produce an answer. To address this, we employ a forced answer generation mechanism: if no answer is provided in the final iteration, the system is instructed to generate an answer based on the existing context.

We implement Self-RAG and CRAG using LangChain⁴ framework. For FLARE and ReAct, we follow their official code implementations. All implementations utilize the same local corpus and retriever as our method for fair comparison. For CRAG and ReAct, we configure DuckDuckGo as the web source, maintaining consistency with PrefRAG.

C DPO Data Construction Details

We randomly sample 15,000 instances from the training set of the 2WikiMQA dataset to construct the training data. First, we use GLM-9B-chat to perform the adaptive retrieval process starting from the q. During the iteration $\tau = \tau_2, \dots, \tau_n$, we configure nine different combinations of model hyperparameter by varying the temperature and top-p values across three different settings $\{0.1, 0.5, 0.9\}$, ensuring a clearer distinction between positive and negative samples. These combinations generate nine predictions during the retrieval source selection process. Each prediction includes a CoT analysis and a status value (i.e., True or False), indicating whether to switch retrieval source. Note that since a single sample generates these predictions across multiple iterations, we also perform random sampling to ensure that the final training samples contain no duplicates and cover data from various iterations. Concurrently, we also use a larger-size parameter model, GLM4-Plus, of the same series to output a gold label for retrieval source selection. In detail, we present the prompt for generating predictions in Table 20 and its input variables in Table 21. The input variables of the prompt together constitute the input x in the training data. Next, we compare the nine predictions generated by GLM4-9B-chat with the gold label. Instances with matching status values form the positive candidate set, while those with differing values form the negative candidate set.

We then use the prompt in Table 22 to compare the data in the positive candidate set with the gold label and employ GLM4-Plus to select the

⁴https://github.com/langchain-ai/langgraph

best instance as the positive sample for training. For the negative sample, we randomly select one instance from the negative candidate set. Additionally, we notice that in the 2WikiMQA dataset, over approximately 70% labels generated by GLM4-Plus have a status value of True. To simulate the real distribution, we select 3,000 instances with a True status value as positive samples y^+ and 1,000 instances with a False status value as negative samples y^- , resulting in 4,000 training samples, denoted as $\{x, y^+, y^-\} \sim \mathcal{D}$.

D Controllable Knowledge Retrieval

We construct two types of controllable retrieval scenarios. Specifically, we collect real-world questions and conduct searches on the open web. The retrieved positive answers are compiled into our corpus. To simulate a more realistic retrieval process, we merge this corpus with the 2WikiMQA corpus to form the final retrieval corpus.

Table 27 presents a controllable response example where the user expects the answer to be generated using knowledge from the local retrieval source. In these cases, specific roles expect the RAG system to rely on knowledge from more controllable local retrieval sources for the final answer while avoiding unfavorable information from the web. Table 28 presents examples where web sources supplement knowledge. Here, specific roles expect the RAG system to supplement local retrieval when its knowledge is insufficient by leveraging web sources. These examples demonstrate that PrefRAG enables users with controllable response needs to prioritize retrieving knowledge from local sources, such as carefully curated brand information. At the same time, it can flexibly incorporate web knowledge when local sources are insufficient. This capability allows RAG to expand its retrieval scope while maintaining control over the retrieval process, thereby improving answer quality and mitigating risks associated with unreliable web information. Consequently, PrefRAG enhances both the adaptability and reliability of RAG systems in real-world applications.

E More Retrieval Sources

In our work, we primarily conduct experiments using two classic retrieval sources with distinct characteristics. However, PrefRAG can support multiple retrieval sources (more than two) along with one predefined retrieval preference in practical

applications. For example, PrefRAG can integrate four retrieval sources, S_1 , S_2 , S_3 , and S_4 , with one designated as the preferred retrieval source, such as S_1 . This requires adjustments to the operations in the two stages of the preference-driven retrieval decision process.

Specifically, in the Retrieve-or-Generate stage, the action space is no longer limited to a single "Search_Engine" action but instead includes four actions: "Search_S1", "Search_S2", "Search_S3", and "Search_S4". The model needs to determine whether to continue retrieval and which source to retrieve based on the existing context. For example, the model determines to continue the retrieval and select "Search_S2" at this stage. We retrieve S_1 following the predefined retrieval preference. If Instruct_Sel determines that a source switch is necessary, we then perform retrieval using " S_2 ".

F Case Study

We conduct a case study, and QA examples of PrefRAG are presented in Table 24, 25, and 26.

In Table 24, given the original query q, "In what year did the Danish plant ecologist who assisted a Danish chemist, famous for the introduction of the concept of pH, die?", PrefRAG first analyzes the q and formulates a reasoning thought: "I need to identify the Danish plant ecologist who assisted a Danish chemist famous for introducing the concept of pH". In iteration τ_1 , PrefRAG retrieves information about the Danish chemist who introduced the concept of pH and identifies him as Søren Peder Lauritz Sørensen. In iteration τ_2 , PrefRAG refines its thought: "Now I need to find the Danish plant ecologist who assisted him". To enhance retrieval accuracy, PrefRAG incorporates the chemist's name into a new subquery: "Danish plant ecologist who assisted Søren Peder Lauritz Sørensen". However, in the next iteration, PrefRAG considers that the retrieval "did not provide specific information about a Danish plant ecologist who assisted Søren Peder Lauritz Sørensen". It then strategizes its goal for the next iteration: "I need to consider if there might be a misunderstanding in the question or if the information is not readily available". At this iteration, the system attempts to generate an answer α , accompanied by a self-reflection label **INCORRECT**, explanation, and improvement suggestions. The self-reflection label correctly identifies that "Not available" is an

incorrect answer. In the explanation and improvement suggestions, the system reflects on the error, noting that the lack of available information on who assisted Søren Peder Lauritz Sørensen prevented it from determining the year of death. It also suggests further historical research or seeking expert consultation in Danish scientific history. A supplementary retrieval is then conducted, which reveals that Carsten Erik Olsen assisted Søren Peder Lauritz Sørensen and provides his birth and death years. With this newly acquired knowledge, the model successfully identifies Carsten Erik Olsen as the Danish plant ecologist in the original query q. In the **Final Answer**, PrefRAG correctly states that Carsten Erik Olsen passed away in 1974 and assigns the self-reflection label as **CORRECT**. The improvement suggestion is: "None needed, the answer is accurate based on the information found".

For comparison, Table 25 presents how ReAct approaches the same question. In some cases, ReAct initially retrieves information from web sources, causing it to miss valuable knowledge from carefully curated local sources. In iteration τ_1 , ReAct correctly identifies that the Danish chemist famous for introducing the concept of pH is Søren Sørensen. However, in iteration τ_2 , it retrieves information from the web suggesting that Thorvald (Thorwald) Julius Sørensen might be connected to the Danish chemist, which is incorrect. Due to this misidentification, ReAct ultimately provides an incorrect year of death for the Danish plant ecologist. By comparing PrefRAG and ReAct, we find that ReAct's initial choice of retrieval sources exhibits a degree of uncertainty. In contrast, PrefRAG follows a preset preference as a guide. Additionally, PrefRAG leverages self-reflection to critically assess its answers, refine subsequent retrieval and reasoning, and generate more reliable and high-quality responses.

Table 26 also presents cases where PrefRAG provided the correct answer on the first attempt. Given the original query q, "Which one was established first, Grouplogic or Inbios?", we observe that PrefRAG follows a clear problem-solving approach: "I need to find the years of establishment of Grouplogic and Inbios to determine which one was established first". It then retrieves "GroupLogic, Inc., founded in 1988" in iteration τ_1 and "InBios International, Inc. was founded in 1996" in iteration τ_2 . Ultimately, PrefRAG correctly identifies **Grouplogic** as the answer, with a self-reflection label of CORRECT.

Settings	HotpotQA	2WikiMQA	MusiQue	BioASQ-Y/N
	Datas	set statistics		
# Samples used for evaluation	500	500	500	500
	Evalue	ation settings		
Metric	Accuracy, F1, EM	Accuracy, F1, EM	Accuracy, F1, EM	Accuracy
	Retrie	eval settings		
Corpus	(Trivedi et al., 2023)	(Trivedi et al., 2023)	(Trivedi et al., 2023)	PubMed
# Documents in Corpus	5233329	139416	430139	23898701
Retriever	BM25, Dense	BM25, Dense	BM25, Dense	BM25, Dense
$top ext{-}k$	3,5,7	3,5,7	3,5,7	3,5,7
	LL	M settings		
# Types of LLMs	5	5	5	5

Table 3: Dataset statistics and experimental settings of different datasets.

M-41-1-0 1 I M-		Hotpe	otQA			2Wik	iMQA			MuS	iQue		BioASQ-Y/N
Methods & LLMs	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
			# Base		thout R		(NoR)	#					
					n-source								
Llama3.1-8B-Instruct	22.6	28.7	23.0	24.8	27.4	30.7	26.4	28.2	3.6	9.4	3.2	5.4	77.8
GLM4-9B-chat	18.4	23.5	17.4	19.8	25.6	29.6	25.0	26.7	3.0	8.8	2.6	4.8	74.0
Llama3.1-70B-Instruct	_32.4_	41.5	31.4	35.1	33.8	37.9	32.6	_34.8_	8.0	14.6	7.4	_10.0	87.0
					prietary								
GPT-4o-mini	29.8	38.4	28.6	32.3	29.2	32.6	26.6	29.5	7.6	15.4	5.0	9.3	86.6
GLM4-Plus	30.2	38.3	29.8	32.8	30.4	35.2	29.6	31.7	8.2	15.8	7.2	10.4	81.8
					a RAG								
					eval sou								
Llama3.1-8B-Instruct	36.4	45.6	34.4	38.8	31.2	35.4	30.2	32.3	6.4	12.2	5.6	8.1	85.8
GLM4-9B-chat	34.8	44.4	34.2	37.8	34.4	38.8	33.8	35.7	8.2	15.0	7.0	10.1	87.2
Llama3.1-70B-Instruct	42.6	53.4	42.6	46.2	45.2	48.2	43.0	45.5	11.4	18.4	10.6	13.5	89.4
GPT-40-mini	45.0	53.8	41.2	46.7	40.2	44.2	38.6	41.0	11.2	19.2	8.8	13.1	89.6
GLM4-Plus	_46.4_	56.7	45.8	49.6	45.6	48.9	_ 43.0	45.8_	15.4	23.5	13.8	_17.6_	89.8
	ncatena										0.0	11.0	00.6
Llama3.1-8B-Instruct	41.6	53.9	41.2	45.6	35.4	39.3	32.6	35.8	9.0	16.0	8.0	11.0	89.6
GLM4-9B-chat	40.8	51.3	39.0	43.7	38.8	43.7	37.4	40.0	9.0	16.7	8.4	11.4	91.0
Llama3.1-70B-Instruct	47.2	59.9	46.8	51.3	49.6	54.0	47.0	50.2	13.4	21.4	12.6	15.8	93.2
GPT-4o-mini	47.4	58.0	44.6	50.0	45.8	49.1	40.6	45.2	13.2	21.3	11.4	15.3	92.2
GLM4-Plus	49.6	61.1	48.4	53.0	48.4	51.7	44.6	48.2	13.6	23.9	13.2	16.9	93.6
ELABE	46.4				ce ARA				166	21.0	14.4	17.6	77.0
FLARE GLM4-Plus	46.4	51.8	41.8	46.7	49.4	45.9	37.8	44.4	16.6	21.9	14.4	17.6	77.2
Self-RAG _{GLM4-Plus}	45.0	54.5	43.6	47.7	32.4	36.7	30.2	33.1	15.4	24.3	13.2	17.6	82.8
CD A C	41.0	50.1			rce RAC			22.0	11.6	17.4	8.8	10.6	89.0
CRAG GLM4-Plus	41.8	50.1	37.8	43.2	35.2	37.6	29.0	33.9	11.6	17.4		12.6	
ReAct w/LR & WR Llama3.1-8B-Instruct	39.4	50.0	37.6	42.3	38.8	39.7	32.0	36.8	13.8	18.4	9.6	13.9	87.2
ReAct w/LR & WR GLM4-9B-chat	44.8 50.2	54.1	40.2 48.8	46.4 53.2	51.6 69.4	51.1 68.4	38.8	47.2	16.0	22.1 33.5	12.6 25.0	16.9 28.4	89.2
ReAct w/ LR & WR Llama3.1-70B-Instruct		60.7			72.2		60.4	66.1	26.6				93.8
ReAct w/LR & WR GPT-4o-mini	51.8	60.3	47.0	53.0		69.9	55.6	65.9	19.0	25.6	14.6	19.7	91.0
ReAct w/LR & WR GLM4-Plus	50.0	59.7 67.0	46.2 53.6	52.0 59.1	64.2 73.8	63.8 70.5	51.8 59.0	59.9 67.8	23.2 25.8	30.6 33.3	18.4 21.2	24.1 26.8	91.8 93.2
$ReAct_{Mix}$ w/ $LR \oplus WR$ $GLM4-Plus$	56.6	67.0	33.0	39.1			39.0	07.8	25.8	33.3	21.2	20.8	93.2
				th Onan	# Ours		inad I I	Me					
ProfP AG	42.0	51.1	700 wii 38.8	n Open 44.0	-source 42.0	ana 1ra 43.2	inea LLi 35.8	40.3	15.4	21.0	12.8	16.4	89.6
PrefRAG Llama3.1-8B-Instruct	45.4	56.3	42.2	44.0	55.0	53.7	42.0	50.2	23.0	29.4	20.0	24.1	87.6
PrefRAG _{GLM4-9B-chat} PrefRAG-DPO _{GLM4-9B-chat}	51.4	57.0	45.0	51.1	57.0	56.0	45.2	52.7	24.2	30.0	20.0	24.1	89.6
ProfP A C	53.6	63.8	51.8	56.4	67.4	66.0	56.8	63.4	27.0	34.3	24.2	28.5	93.2
PrefRAG Llama3.1-70B-Instruct	_33.0_	_03.8 _			h Propri			_05.4_	_27.0			_20.3_	93.4
PrefRAG GPT-40-mini	58.6	66.0	50.4	urs wiii 56.6	n Propri 76.2	етагу L1 72.1	.м.s 59.4	69.2	28.2	34.3	21.2	27.9	92.8
	13.6↑	12.2↑	9.2↑	9.9↑	36.0↑	72.1 27.9↑	20.8↑	28.2↑	26.2	34.3 15.1↑	12.4	27.9 14.8↑	3.2↑
Δ GPT-4o-mini \rightarrow Vanilla w/LR	11.2	8.0↑	9.2↑ 5.8↑	9.9↑	30.4↑	23.0↑	18.8↑	24.1	15.0	13.1	9.8↑	12.6	0.6↑
$\Delta_{\text{GPT-4o-mini} \rightarrow \text{Vanilla}_{Mix}} \text{ w/LR} \oplus \text{WR}$ PrefRAG _{GLM4-Plus}	59.0	68.4	55.0	60.8	79.6	76.7	65.2	73.8	32.2	39.4	27.4	33.0	94.0
	12.6↑	11.7	9.2↑	11.2↑	7 9.0 34.0↑	7 0. 7 27.8↑	22.2↑	28.0↑	16.8↑	39.4 15.9↑	13.6↑	33.0 15.4↑	9 4. 0 4.2↑
Δ GLM4-Plus \rightarrow Vanilla w/LR	9.4↑	7.3↑	9.2↑ 6.6↑	7.8↑	31.2↑	25.0↑	20.6	25.6↑	18.6	15.5↑	14.2	16.1	4.2↑
Δ GLM4-Plus \rightarrow Vanilla _{Mix} w/LR \oplus WR	9.4	1.5	0.0	7.0	31.4	25.0	20.0	23.0	10.0	13.3	14.4	10.1	0.4

Table 4: Results (%) of overall performance on all models and datasets.

Retrieval Sources		Hotp	otQA			2Wik	iMQA			Mus	iQue		BioASQ-Y/N
Retrieval Sources	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
					Llama	3.1-8B-	Instruci	t					
Local Retrieval (LR)	36.4	45.6	34.4	38.8	31.2	35.4	$-30.\overline{2}$	$-3\overline{2}.\overline{3}$	6.4	$^{-1}\bar{2}.\bar{2}^{-}$	5.6	8.1	85.8
Web Retrieval (WR)	36.0	45.1	34.2	38.4	28.6	31.4	24.0	28.0	5.4	10.8	4.6	6.9	87.6
$LR \oplus WR$	41.6	53.9	41.2	45.6	35.4	39.3	32.6	35.8	9.0	16.0	8.0	11.0	89.6
					GLi	M4-9B-	chat						
Local Retrieval (LR)	34.8	44.4	34.2	37.8	34.4	38.8	$-3\bar{3}.\bar{8}$	$^{-}3\bar{5}.\bar{7}$	8.2	$^{-1}\bar{5}.\bar{0}$	7.0	10.1	87.2
Web Retrieval (WR)	39.4	48.0	35.8	41.1	34.4	38.8	32.4	35.2	5.6	12.8	4.8	7.7	87.4
$LR \oplus WR$	40.8	51.3	39.0	43.7	38.8	43.7	37.4	40.0	9.0	16.7	8.4	11.4	91.0
					Llama3	.1-70B	-Instruc	rt .					
Local Retrieval (LR)	42.6	53.4	42.6	46.2	45.2	$-48.\overline{2}$	$-4\overline{3}.\overline{0}$	$^{-}4\overline{5}.\overline{5}$	$^{-}1\overline{1}.\overline{4}$	$^{-18.4}$	10.6	13.5	89.4
Web Retrieval (WR)	38.8	50.0	38.4	42.4	36.6	38.3	30.4	35.1	9.4	15.4	8.8	11.2	89.6
$LR \oplus WR$	47.2	59.9	46.8	51.3	49.6	54.0	47.0	50.2	13.4	21.4	12.6	15.8	93.2
					GI	PT-40-n	nini						
Local Retrieval (LR)	45.0	53.8	41.2	46.7	40.2	$-44.\overline{2}$	$^{-}38.\overline{6}$	$-41.\overline{0}$	$^{-}1\overline{1}.\overline{2}$	19.2	8.8	13.1	89.6
Web Retrieval (WR)	43.4	53.4	41.0	45.9	34.4	39.8	31.0	35.1	10.2	17.7	9.2	12.4	90.2
$LR \oplus WR$	47.4	58.0	44.6	50.0	45.8	49.1	40.6	45.2	13.2	21.3	11.4	15.3	92.2
					G.	LM4-P	lus						
Local Retrieval (LR)	46.4	56.7	45.8	49.6	45.6	48.9	$-4\bar{3}.\bar{0}$	$-45.\overline{8}$	⁻ 1 5 .4	$^{-}2\overline{3}.5^{-}$	13.8	17.6	89.8
Web Retrieval (WR)	45.8	55.6	43.4	48.3	39.2	42.9	36.2	39.4	11.4	18.6	10.6	13.5	91.8
$LR \oplus WR$	49.6	61.1	48.4	53.0	48.4	51.7	44.6	48.2	13.6	23.9	13.2	16.9	93.6

Table 5: All results (%) of Vanilla with different retrieval sources.

G4 . 4 . •		Hotp	otQA			2Wik	iMQA			Mus	iQue		BioASQ-Y/N
Strategies	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM Avg.		Acc.
					Lla	ma-3.1-	8B-Instr	uct					
PrefRAG Direct	40.6	48.0	37.0	41.9	⁻ 38.8	$^{-}4\overline{0}.\overline{1}$	$^{-}3\overline{2}.\overline{0}$	$^{-}37.\overline{0}$	$-12.\overline{2}$	_ 1 7.8	10.2	_ <u>1</u> 3.4 _	87.0
PrefRAG	42.0	51.1	38.8	44.0	42.0	43.2	35.8	40.3	15.4	21.0	12.8	16.4	89.6
						GLM4-9	9B-chat						
PrefRAG Direct	45.2	50.8	37.8	44.6	51.2	⁻ 4 9 .9	$^{-}38.\overline{6}$	$-46.\overline{6}$	- 1 4. 8	21.6	12.4		8 8.8
PrefRAG	45.4	56.3	42.2	48.0	55.0	53.7	42.0	50.2	23.0	29.4	20.0	24.1	87.6
					Llan	na-3.1-7	OB-Inst	ruct					
PrefRAG Direct	50.4	62.4	49.6	54.1	61.4	$^{-}6\overline{2}.\overline{5}$	$^{-}5\overline{4}.\overline{8}$	⁻ 5 9 . 6	$-\bar{23.0}$	<u></u>	21.2	⁻ 24.7 ⁻	93.4
PrefRAG	53.6	63.8	51.8	56.4	67.4	66.0	56.8	63.4	27.0	34.3	24.2	28.5	93.2
						GPT-4	o-mini						
PrefRAG Direct	55.8	63.3	49.4	56.2	76.2	71.9	⁻ 5 9 .4	$^{-}6\bar{9}.\bar{2}$	$-\bar{28.4}$	34.3	20.8	77.8	92.4
PrefRAG	58.0	66.0	50.4	58.3	76.2	72.1	59.4	69.2	28.2	34.3	21.2	27.9	92.8
						GLM4	1-Plus						
PrefRAG Direct	56.4	66.3	52.4	58.4	$^{-}7\overline{5}.6$	$^{-}7\overline{2}.\overline{2}$	$^{-}6\overline{0}.\overline{6}$	$^{-}6\bar{9}.\bar{5}$	$-\bar{29.8}$	36.1	24.6	30.2	92.6
PrefRAG	59.0	68.4	55.0	60.8	79.6	76.7	65.2	73.8	32.2	39.4	27.4	33.0	94.0

Table 6: Results (%) of different PrefRAG strategies.

Methods		Hotp	otQA (Count))	
Westons	Total Local Num	Total Web Num	Total Num	Used Local Num	Used Num
		Llama3.1-8B-Instr	ruct		
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	₁₃₄₀	₁₃₄₀	2680	₁₃₄₀	
PrefRAG	1025	347	1372	736	1083
Δ Retrieval Counts	315↓	993↓	1308↓	604↓	1597 ↓
		GLM4-9B-chai	ţ		
ReAct w/ LR ⊕ WR	₁₁₁₀	₁₁₁₀	$ 2\overline{2}2\overline{0}$ $ -$	₁₁₁₀	$-2\bar{2}2\bar{0}$
PrefRAG	1274	480	1754	794	1274
Δ Retrieval Counts	164↑	630↓	466↓	316↓	946↓
PrefRAG+DPO	1308	579	1887	729	1308
Δ Retrieval Counts	198↑	531↓	333↓	381↓	912↓
		Llama3.1-70B-Inst	ruct		
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	930	- 9 3 0	1860	930	1860
PrefRAG	1025	289	1314	736	1025
Δ Retrieval Counts	95↑	641↓	546↓	194↓	835↓
		GPT-40-mini			
ReAct w/ LR ⊕ WR	1040	1040	2080	1040	2080
PrefRAG	1113	371	1484	742	1113
Δ Retrieval Counts	73↑	669↓	596↓	298↓	967 ↓
		GLM4-Plus			
ReAct w/ LR ⊕ WR	7 94	7 9 4	$^{-}$ $^{-}$ $^{1}\overline{5}8\overline{8}$ $^{-}$ $^{-}$	7 94	1588
PrefRAG	1031	248	1279	783	1031
Δ Retrieval Counts	237↑	546↓	309↓	11↓	557↓

Table 7: Total retrieval counts on HotpotQA dataset.

			Но	tpotQA							
Methods	Perfor	mance (%) (†)	Counts of F	Retrieval (\dagger)						
	Acc. F1 EM			Total Num	Used Num						
	Llan	1a3.1-8E	3-Instruc	rt							
ReAct w/ LR WR	-41.8	52.0	39.0		2680						
PrefRAG	42.0	51.1	38.8	1372	1083						
	(<i>ELM4-91</i>	3-chat								
ReAct w/ LR WR	48.4	56.0	42.6		2220						
PrefRAG	45.4	56.3	42.2	1754	1274						
PrefRAG+DPO	51.4	57.0	45.0	1887	1308						
	Llam	a3.1-701	B-Instru	ct							
ReAct w/ LR WR	51.6	63.7	50.6	1860	1860						
PrefRAG	53.6	63.8	51.8	1314	1025						
		GPT-40-	mini								
ReAct w/ LR WR	- 5 7.0	65.9	51.4		2080						
PrefRAG	58.6	66.0	50.4	1484	1113						
	GLM4-Plus										
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	56.6	67.0	53.6	1588	1588						
PrefRAG	59.0	68.4	55.0	1279	1031						

Table 8: Efficiency and accuracy trade-off on HotpotQA dataset.

Mala		2Wik	iMQA (Count	i)	
Methods	Total Local Num	Total Web Num	Total Num	Used Local Num	Used Num
		Llama3.1-8B-Insti	ruct		
ReAct w/ LR ⊕ WR	$ \overline{1207}$ $ -$	₁₂₀₇	$-2\overline{4}1\overline{4}$	$ 1\overline{2}0\overline{7}$	2414 -
PrefRAG	1134	513	1647	623	1136
Δ Retrieval Counts	73↓	694↓	767 ↓	584↓	1278 ↓
		GLM4-9B-char	t .		
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	₁₂₅₉	₁₂₅₉	-2518	1259	- ⁻ 2 5 1 8 -
PrefRAG	1330	387	1717	943	1330
Δ Retrieval Counts	71↑	872↓	801↓	316↓	1188↓
PrefRAG+DPO	1354	431	1785	923	1354
Δ Retrieval Counts	95↑	828↓	733 ↓	336↓	1164↓
		Llama3.1-70B-Inst	ruct		
ReAct w/ LR ⊕ WR	₁₁₈₉	₁₁₈₉	$-2\overline{3}78$	1189	
PrefRAG	1132	271	1403	861	1132
Δ Retrieval Counts	57↓	918↓	975↓	328↓	1246 ↓
		GPT-40-mini			
ReAct w/ LR ⊕ WR	1302	1302	2604	1302	2604
PrefRAG	1357	485	1842	872	1357
Δ Retrieval Counts	55↑	817↓	762 ↓	430↓	1247 ↓
		GLM4-Plus			
ReAct w/ LR ⊕ WR	913	9 1 3	- ₁₈₂₆	913	- ₁₈₂₆ -
PrefRAG	1200	248	1448	952	1200
Δ Retrieval Counts	287↑	665↓	378↓	39↑	626↓

Table 9: Total retrieval counts on 2WikiMQA dataset.

			2Wi	ikiMQA						
Methods	Perfor	mance (%) (†)	Counts of F	Retrieval (\lambda)					
	Acc. F1 EM			Total Num	Used Num					
	Llan	1a3.1-8E	3-Instruc	rt						
ReAct w/ LR WR	38.0	39.4	30.6	2414	2414					
PrefRAG	42.0	43.2	35.8	1647	1136					
	(<i>ELM4-91</i>	3-chat							
ReAct w/ LR WR	56.8	54.6	41.2	2518	2518					
PrefRAG	55.0	53.7	42.0	1717	1330					
PrefRAG+DPO	57.0	56.0	45.2	1785	1354					
	Llam	a3.1-70i	B-Instru	ct						
ReAct w/ LR ⊕ WR	68.2	68.7	61.4	2378	2378					
PrefRAG	67.4	66.0	56.8	1403	1132					
		GPT-40-	mini							
ReAct w/ LR WR	78.4	74.1	61.8	2604	2604					
PrefRAG	76.2	72.1	59.4	1842	1357					
	GLM4-Plus									
$\overline{ReAct} \ \overline{w} / \overline{LR} \oplus \overline{WR}$	73.8	70.5	59.0	1826	1826					
PrefRAG	79.6	76.7	65.2	1448	1200					

Table 10: Efficiency and accuracy trade-off on 2WikiMQA dataset.

Methods		Mus	siQue (Count)		
Wethous	Total Local Num	Total Web Num	Total Num	Used Local Num	Used Num
		Llama3.1-8B-Insti	ruct		
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	$ \frac{1}{1444}$	1 444	-2888	1444	2888
PrefRAG	1369	695	2064	675	1370
Δ Retrieval Counts	75↓	749↓	824↓	769 ↓	1518↓
		GLM4-9B-char	t		
ReAct w/ LR WR	1 4 7 8	₁₄₇₈	2956	₁₄₇₈	2956
PrefRAG	1625	835	2460	790	1625
Δ Retrieval Counts	147↑	643↓	496↓	688↓	1331↓
PrefRAG+DPO	1643	996	2639	647	1643
Δ Retrieval Counts	165↑	482↓	317↓	831↓	1313↓
		Llama3.1-70B-Inst	truct		
ReAct w/ LR ⊕ WR	$ \frac{1}{1241}$	1 2 4 1	$-2\overline{4}8\overline{2}$	1241	2482
PrefRAG	1170	452	1622	718	1170
Δ Retrieval Counts	71↓	789↓	860↓	523↓	1312↓
		GPT-40-mini			
ReAct w/ LR ⊕ WR	1515	1515	3030	1515	3030
PrefRAG	1562	885	2447	677	1562
Δ Retrieval Counts	47↑	630↓	583↓	838↓	1468↓
		GLM4-Plus			
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	918	9 1 8	1836	918	1836
PrefRAG	1373	603	1976	770	1373
Δ Retrieval Counts	455↑	315↓	140↑	148↓	463↓

Table 11: Total retrieval counts on MusiQue dataset.

	MusiQue										
Methods	Performance (%) (†)			Counts of F	Retrieval (\(\psi \)						
	Acc.	F1	EM	Total Num	Used Num						
	Llan	na3.1-8B	-Instruc	t							
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	- _{12.8}	_ <u>1</u> 9.3 _	10.4	- $ 2888$ $ -$	2888						
PrefRAG	15.4	21.0	12.8	2064	1370						
	(3LM4-9E	3-chat								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	$-2\bar{2}.\bar{0}$	28.7	18.8								
PrefRAG	23.0	29.4	20.0	2460	1625						
PrefRAG+DPO	24.2	30.0	20.2	2639	1643						
	Llam	a3.1-701	B-Instruc	rt .							
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	-25.0	34.0	23.8	₂₄₈₂							
PrefRAG	27.0	34.3	24.2	1622	1170						
		GPT-40-	mini								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	$^{-}28.\overline{4}$	34.8	21.4	3030	3030						
PrefRAG	28.2	34.3	21.2	2447	1562						
		GLM4-	Plus								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	-25.8	33.3	21.2	₁₈₃₆	1836						
PrefRAG	32.2	39.4	27.4	1976	1373						

Table 12: Efficiency and accuracy trade-off on MusiQue dataset.

Methods	BioASQ-Y/N (Count)									
Wethous	Total Local Num	Total Web Num	Total Num	Used Local Num	Used Num					
		Llama3.1-8B-Insti	ruct							
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	$ \overline{1205}$ $ -$	₁₂₀₅	$-2\overline{4}1\overline{0}$	-1205	2410					
PrefRAG	1178	210	1388	968	1178					
Δ Retrieval Counts	27 ↓	995↓	1022↓	237↓	1232↓					
		GLM4-9B-cha	t							
ReAct w/ LR WR	829	8 2 9	1658		1658					
PrefRAG	1116	213	1329	934	1147					
Δ Retrieval Counts	287↑	616↓	329 ↓	105↑	511↓					
PrefRAG+DPO	1383	726	2109	657	1383					
Δ Retrieval Counts	554↑	103↓	451↑	172↓	275↓					
		Llama3.1-70B-Inst	truct							
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	1 0 9 7	₁₀₉₇		₁₀₉₇	2194					
PrefRAG	1052	472	1524	666	1138					
Δ Retrieval Counts	45↓	625↓	670↓	431↓	1056↓					
		GPT-40-mini								
ReAct w/ LR ⊕ WR	714	714	1428	714	1428					
PrefRAG	799	202	1001	599	801					
Δ Retrieval Counts	85↑	512↓	427 ↓	115↓	627 ↓					
		GLM4-Plus								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	665	665	1330		1330					
PrefRAG	681	118	799	583	701					
Δ Retrieval Counts	16↑	547↓	531↓	82↓	629↓					

Table 13: Total retrieval counts on BioASQ-Y/N dataset.

	BioASQ-Y/N								
Methods	Performance (%) (†)	Counts of F	Retrieval (\dagger)						
	Acc.	Total Num	Used Num						
	Llama3.1-8B-Instruc	et .							
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	87.8	$-2\overline{4}1\overline{0}$	2410						
PrefRAG	89.6	1388	1178						
	GLM4-9B-chat								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	8 7 . 4	1658	1658						
PrefRAG	87.6	1329	1147 1383						
PrefRAG+DPO	89.6	2109							
	Llama3.1-70B-Instru	ct							
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	9 3.6	2194	2194						
PrefRAG	93.2	1524	1138						
	GPT-40-mini								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	9 1.4	-1428	1428						
PrefRAG	92.8	1001	801						
	GLM4-Plus								
$\overline{ReAct} \overline{w} / \overline{LR} \oplus \overline{WR}$	$\overline{93.2}$	1330	1330						
PrefRAG	94.0	799	701						

Table 14: Efficiency and accuracy trade-off on BioASQ-Y/N dataset.

LLMs	Methods	HotpotQA			2WikiMQA			MusiQue				BioASQ-Y/N		
LLWIS	Methods	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
	PrefRAG	42.0	51.1	38.8	44.0	42.0	43.2	35.8	40.3	15.4	21.0	12.8	16.4	89.6
Llama3.1-8B-Instruct	w/o Pref-AR	41.0	<u>50.9</u>	39.8	43.9	36.0	37.8	30.2	34.7	13.6	19.0	11.0	14.5	<u>81.4</u>
	w/o Self-Reflection	<u>41.6</u>	<u>50.9</u>	<u>39.6</u>	44.0	<u>41.6</u>	<u>42.1</u>	<u>34.4</u>	<u>39.4</u>	13.2	<u>19.9</u>	12.2	15.1	89.6
	PrefRAG	45.4	56.3	42.2	48.0	55.0	53.7	42.0	50.2	23.0	29.4	20.0	24.1	87.6
GLM4-9B-chat	w/o Pref-AR	46.8	54.8	<u>42.2</u>	47.9	51.0	51.2	39.0	47.1	16.0	22.5	12.8	17.1	87.0
	w/o Self-Reflection	47.0	57.4	45.0	49.8	53.8	54.3	43.4	50.5	21.4	27.4	18.2	22.3	89.0
	PrefRAG	51.4	57.0	45.0	51.1	57.0	56.0	45.2	52.7	24.2	30.0	20.2	24.8	89.6
GLM4-9B-chat-DPO	w/o Pref-AR	47.4	53.4	41.0	47.3	53.6	53.4	40.0	49.0	18.0	23.1	14.4	18.5	88.8
	w/o Self-Reflection	<u>49.4</u>	<u>56.0</u>	<u>42.6</u>	49.3	56.8	<u>54.4</u>	<u>41.8</u>	<u>51.0</u>	22.4	28.0	18.4	22.9	89.8
	PrefRAG	53.6	63.8	51.8	56.4	67.4	66.0	56.8	63.4	27.0	34.3	24.2	28.5	93.2
Llama3.1-70B-Instruct	w/o Pref-AR	<u>52.6</u>	<u>63.5</u>	<u>51.4</u>	55.8	64.8	63.8	54.8	61.1	25.4	33.6	22.4	27.1	<u>92.4</u>
	w/o Self-Reflection	51.4	63.0	49.4	54.6	66.2	66.1	57.0	<u>63.1</u>	26.8	34.2	24.4	28.5	92.2
	PrefRAG	58.6	66.0	50.4	58.3	76.2	72.1	<u>59.4</u>	69.2	28.2	34.3	21.2	27.9	92.8
GPT-4o-mini	w/o Pref-AR	51.4	58.4	43.8	51.2	69.8	66.8	52.6	63.1	19.6	26.7	14.4	20.2	89.4
	w/o Self-Reflection	57.8	66.2	51.6	58.5	76.6	71.9	59.8	69.4	28.6	33.7	21.0	27.8	92.4
	PrefRAG	59.0	68.4	55.0	60.8	79.6	76.7	65.2	73.8	32.2	39.4	27.4	33.0	94.0
GLM4-Plus	w/o Pref-AR	51.6	61.1	47.8	53.5	74.2	72.6	59.6	68.8	26.2	33.3	22.0	27.2	93.4
	w/o Self-Reflection	57.6	67.3	53.8	59.6	<u>78.6</u>	74.8	62.8	72.1	32.0	38.5	27.0	32.5	93.6

Table 15: All results (%) of ablation study.

top-k	top-k Methods		HotpotQA			2WikiMQA			MusiQue				BioASQ-Y/N	
F		Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
top-3	Vanilla RAG w/ LR ⊕ WR	47.8	58.5	45.8	50.7	46.4	50.6	43.6	46.9	13.8	23.4	13.2	16.8	92.8
	PrefRAG	56.2	66.6	52.6	58.5	79.6	75.9	64.6	73.4	30.6	38.2	26.8	31.9	93.4
top-5	Vanilla RAG w/ LR ⊕ WR	49.6	61.1	48.4	53.0	48.4	51.7	44.6	48.2	13.6	23.9	13.2	16.9	93.6
	PrefRAG	59.0	68.4	55.0	60.8	79.6	76.7	65.2	73.8	32.2	39.4	27.4	33.0	94.0
top-7	Vanilla RAG w/ LR ⊕ WR	49.6	61.1	48.6	53.1	49.6	53.5	45.8	49.6	15.4	24.3	13.6	17.8	93.4
	PrefRAG	58.2	68.4	54.8	60.5	81.0	77.3	65.8	74.7	32.2	39.4	28.6	33.4	93.0

Table 16: Results (%) of different top-k values on the GLM4-Plus model.

Dotniovou	HotpotQA			2WikiMQA			MusiQue				BioASQ-Y/N		
Retriever	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.	F1	EM	Avg.	Acc.
PrefRAG (bge-large-en-v1.5)	59.8	68.9	56.0	61.6	75.4	72.4	62.0	69.9	31.8	39.6	28.4	33.3	91.6
PrefRAG (BM25)	59.0	68.4	55.0	60.8	79.6	76.7	65.2	73.8	32.2	39.4	27.4	33.0	94.0

Table 17: Results (%) of different retrievers on the GLM4-Plus model.

Overall Prompt

Instructions

Answer the following questions as best you can. When you need to search more information, You have access to the following tools:

{tool}

Question: the input question you must answer

Use the following format for each step:

Thought: you should always think about what to do

Action: the action to take, should be one of {tool_name} if it needed (Make sure to use

the exact tool name from the list).

Action Input: the input of the action

Observation: the result of the action

... (this Thought/Action/Action Input/Observation should not be repeated more than {max_step} times. If it exceeds {max_step} times, the final answer should be given directly.)

Thought: I now know the final answer to the original question

Final Answer: {answer_format}

After providing the Final Answer, evaluate the response:

Self-Evaluation: Describe the accuracy of the Final Answer by choosing one of [CORRECT[CORRECT]/PARTIALLY CORRECT[PARTIALLY CORRECT]/INCORRECT[INCORRECT]].

Explanation: Briefly explain why you chose the label.

Improvement Suggestions: Optionally suggest how the answer could be improved if needed (omit this if the answer is correct).

Note: Ensure the Final Answer strictly follows the format: {answer_format}

Begin!

Question: {question}

{thought}

Table 18: Overall prompt for PrefRAG.

Details of Input Variables in the Overall Prompt

{tool}

The tool represents the retrieval tool, and its details are as follows.

```
Search_Engine:
{
    "name": "Search_Engine",
    "description": "This is a knowledge base general search engine that can be used to
        query external knowledge, learn facts, etc.",
    "input": "The phrase or question to be searched."
}
```

{tool_name}

The name of the retrieval tool.

{max_step}

The {max_step} defines the threshold for the number of iterations of "Thought/Action/Action Input/Observation" in the overall prompt, acting as a soft limit. Given the potentially limited instruction-following ability of some LLMs, we have also implemented a hard threshold in our method, set to {max_step}+1.

{answer_format}

For multi-hop dataset:

Provide the most concise answer to the original input question. Give me only the final answer without including any other words.

For multi-choice dataset:

Provide the correct option to the original question. Answer with only the letter (e.g., A, B, ...) without including any other words.

{question}

Original question.

{thought}

The {thought} encompasses all the reasoning processes that have occurred so far, including *Thought*, *Action Input*, and *Observation*. Initially, {thought} contains no content.

Table 19: Input variables in the overall prompt for PrefRAG.

Preference-Driven Retrieval Source Selection Prompt

Instructions

You are tasked with evaluating whether newly retrieved information provides additional insights or value for answering an original question. Follow these steps carefully:

Steps:

- **1.** Compare the new information (labeled as "New information") against the information already obtained (labeled as "Information already obtained").
- **2.** Determine if the "New information" includes any new details or elements that directly contribute to solving the "Original question."
- **3.** Output the result as a dictionary in the following JSON format: json {{ "analysis": "<your analysis here>", "status": "<True or False>" }}

Key points:

- "New information" is considered helpful if it adds relevant details that were previously absent and assists in answering the original question.
- Irrelevant, redundant, or already-known information should result in "status": "False".

Original question: {question}

Information already obtained: {existed_info}

New information: {observation}

Your task: Judging based on the above content, has new, useful information been pro-

vided?

Table 20: Preference-driven retrieval source selection prompt for PrefRAG.

Details of Input Variables in the Preference-Driven Retrieval Source Selection Prompt

{question}

Original question.

{existed_info}

The existed_info refers to all the valid passages retrieved in previous iterations up to this point, which had already been organized within the *Observation* during those iterations.

{observation}

The {observation} refers to the top - k passages retrieved during the current iteration.

Table 21: Input variables in the preference-driven retrieval source selection prompt for PrefRAG.

Prompt for Obtaining the Positive Sample of Retrieval Source Selection Stage

Instructions

List to compare: {analysis}.

I will provide you with a standard answer analysis. Compare the standard answer analysis with the results in the list below to determine which one is the most similar.

```
Output the result as a dictionary in the following JSON format:
json {{ "id": "<entry_id of the most similar analysis>"}}

Standard answer analysis: {label_analysis}.
```

Table 22: The prompt for obtaining the positive sample of retrieval source selection stage to construct training data.

Details of Input Variables in the Prompt for Obtaining Preferred Retrieval Labels

{label_analysis}

The {label_analysis} represents the output generated by the GLM-4-Plus LLM in response to the preference selection prompt.

{analysis}

The {analysis} represents the output from the smaller-size parameter LLM in response to the preference selection prompt and takes the form of a list containing multiple reasoning results.

Table 23: Input Variables in prompt for obtaining preferred retrieval labels in the training data.

Question: In what year did the Danish plant ecologist who assisted a Danish chemist, famous for the introduction of the concept of pH die?

[Thought: I need to identify the Danish plant ecologist who assisted a Danish chemist famous for introducing the concept of pH.]

Action: Search_Engine Action Input: "Danish chemist who introduced the concept of pH"

Observation: . . . top-k retrieved passages . . . <u>title</u>: S. P. L. Sørensen+ <u>content</u>: Søren Peder Lauritz Sørensen (9 January 1868 – 12 February 1939) was a Danish chemist, famous for the introduction of the concept of pH, a scale for measuring acidity and alkalinity . . .

Thought: The Danish chemist who introduced the concept of pH is Søren Peder Lauritz Sørensen. Now I need to find the Danish plant ecologist who assisted him.

Action: Search_Engine Action Input: "Danish plant ecologist who assisted Søren Peder Lauritz Sørensen"

Observation: ... top-k retrieved passages ...

Thought: The search did not yield specific information about a Danish plant ecologist who assisted Søren Peder Lauritz Sørensen. I need to refine my search to find the specific plant ecologist.

Observation: . . . top-k retrieved passages . . . title: S. P. L. Sørensen - Wikipedia content: Søren Peter Lauritz Sørensen . . . was a Danish chemist, known for the introduction of the concept of pH . . . title: Søren Peder Lauritz Sørensen (January 9, 1868 - Prabook content: Søren Peder Lauritz Sørensen was a Danish chemist, famous for the introduction of the concept of pH . . . title: Søren Peter Lauritz Sørensen - Encyclopedia.com content: Søren Peter Lauritz Sørensen . . . for introducing the pH . . .

Thought: The search did not provide specific information about a Danish plant ecologist who assisted Søren Peder Lauritz Sørensen. I need to consider if there might be a misunderstanding in the question or if the information is not readily available.

Final Answer: Not available X Self-Evaluation: INCORRECT

Explanation: The search did not yield specific information about a Danish plant ecologist who assisted Søren Peder Lauritz Sørensen, making it impossible to provide the year of death for such an individual.

Improvement Suggestions: Further detailed historical research or access to specific academic papers might be required to find the exact individual and their year of death. Alternatively, rephrasing the question or seeking expert consultation in Danish scientific history could provide more accurate results.

Thought: I have identified Carsten Erik Olsen as the Danish plant ecologist who assisted Sren Peder Lauritz Srensen.

Final Answer: 1974 ✓ Self-Evaluation: CORRECT

Explanation: The search correctly identified Carsten Erik Olsen as the Danish plant ecologist who assisted Søren Peder Lauritz Sørensen, and his year of death is 1974.

Improvement Suggestions: None needed, the answer is accurate based on the information found.

Golden Answer: 1974

Table 24: An example of PreRAG on HotpotQA dataset (Self-Evaluation: INCORRECT). "Bold blue text" represents all segments related to solving the original query (including both relevant and correct segments as well as relevant but incorrect segments); green segments denote those used to support correct answers; red segments represent key incorrect information and erroneous responses collected; the "\scriv" indicates a correct answer, while "\scriv" signifies an incorrect answer.

Question: In what year did the Danish plant ecologist who assisted a Danish chemist, famous for the introduction of the concept of pH die?

[Thought: I need to identify the Danish plant ecologist who assisted a Danish chemist famous for introducing the concept of pH.]

Action: Search_Web Action Input: "Danish chemist who introduced the concept of pH"

Observation: ... top-k retrieved passages ...

<u>title</u>: S. P. L. Sørensen - Wikipedia <u>content</u>: From 1901 to 1938, <mark>Sørensen . . . he introduced the pH-scale as a simple way of expressing it in 1909 . . .</mark>

title: Søren Sørensen - Science History Institute of pH as a convenient way of expressing acidity . . .

<u>title</u>: Søren Sørenson: The Pioneer of pH - Proto Magazine <u>content</u>: Sørensen's invention was the pH scale. (The term pH refers simply to "the power of hydrogen" . . .

title: Today in Chemistry History: Søren Sørensen and the pH scale content: Click to enlarge On this day (9th January) in 1868, Danish chemist Søren Sørensen was born. He's best known for developing the pH scale, which . . .

<u>title</u>: The Concept of pH and its Definition - Unacademy <u>content</u>: The concept of pH was introduced first by Søren Sørensen, a Danish Chemist, in 1909. It . . .

Thought: The Danish chemist famous for introducing the concept of pH is Søren Sørensen. Now I need to find the Danish plant ecologist who assisted him.

Action: Search_Web Action Input: "Danish plant ecologist who assisted Søren Sørensen"

Observation: ... top-k retrieved passages ...

<u>title</u>: Thorvald Sørensen - Wikipedia <u>content</u>: He developed a quotient of similarity in species composition between plant communities-the still much-used Sørensen similarity index . . .

<u>title</u>: S. P. L. Sørensen - Wikipedia <u>content</u>: Sørensen was born in Havrebjerg Denmark in 1868 as the son of a farmer. He began his studies at the University of Copenhagen . . .

<u>title</u>: Søren Johannes Sørensen - Københavns Universitets Forskningsportal <u>content</u>: Søren Johannes Sørensen. . . . These are related to our in vitro model systems, plant and animal models to field and cohord studies, . . .

<u>title</u>: Famous Scientists from Denmark | List of Top Danish Scientists - Ranker <u>content</u>: Thorvald (Thorwald) Julius Sørensen (4 July 1902 - 21 June 1973) was a Danish botanist and evolutionary biologist. Sørensen was professor at the Royal Veterinary and Agricultural University 1953-1955 and at the University of Copenhagen 1955-1972. He was director of the Copenhagen Botanical Garden and Botanical Museum during the same period.

title: Søren J. Sørensen content: Lars Hestbjerg Hansen Department of Plant and Environmental Science . . .

Thought: The Danish plant ecologist who assisted Søren Sørensen is **Thorvald Sørensen**. I now know the final answer to the original question

Final Answer: 1973 X

Golden Answer: 1974

Table 25: An example of ReAct (w/ LR & WR) on HotpotQA dataset. Bold blue text represents all segments related to solving the original query (including both relevant and correct segments as well as relevant but incorrect segments). Green segments denote those used to support correct answers. Red segments represent key incorrect information and erroneous responses collected. The "\scriv" indicates a correct answer, while "\scriv" signifies an incorrect answer.

Question: Which one was established first, Grouplogic or Inbios?

[Thought: I need to find the establishment years of Grouplogic and Inbios to determine which one was established first.]

Action: Search_Engine Action Input: "Grouplogic establishment year"

Observation: ... top-k retrieved passages ...

title: GroupLogic+ content: GroupLogic, Inc., founded in 1988 and headquartered in Arlington, Virginia, USA, is an enterprise software company that develops, sells and supports software for moving and storing data including activEcho, mobilEcho, ArchiveConnect, MassTransit and ExtremeZ- IP. GroupLogic's products are used by information technology organizations to allow employees to access and manage corporate files regardless of the type of computing platform the employee is using to access the network. On September 13, 2012, GroupLogic announced that it became a subsidiary of Acronis, a software company specializing in backup and disaster recovery products and services.

title: Indian Ecclesiastical Establishment+ content: The Indian Ecclesiastical Establishment was ...

title: Up the Establishment+ content: Up the Establishment is a 1969 . . .

title: List of schools in Bangalore+ content: This page provides a list of schools in Bangalore, India....

title: Tommys+ content: Tommys is a dansband from Vaasa, established in 1980. . . .

Thought: I found that Grouplogic was established in 1988. Now I need to find the establishment year of Inbios. Action: Search_Engine Action Input: "Inbios establishment year"

Observation: ... top-k retrieved passages ...

title: InBios+ content: InBios International, Inc. is a medical diagnostic company based in Seattle that specializes in the detection of infectious diseases. The company was founded in 1996, and since its inception has developed several technologies useful in designing rapid and ELISA based immunodiagnostic assays ...

title: Indian Ecclesiastical Establishment+ content: The Indian Ecclesiastical Establishment was created in 1813 . . .

title: Up the Establishment+ content: Up the Establishment is a 1969 West German comedy film . . .

<u>title</u>: List of schools in Bangalore+ <u>content</u>: This page provides a list of schools . . . in Bangalore

title: Tommys+ content: Tommys is a dansband from Vaasa, established in 1980 . . .

Thought: I now know the final answer to the original question

Final Answer: Grouplogic ✓ Self-Evaluation: CORRECT

Explanation: Grouplogic was established in 1988, while Inbios was established in 1996. Therefore, Grouplogic was established first. The answer is correct and concise. No improvement is needed.

Golden Answer: Grouplogic

Table 26: An example of PreRAG on 2WikiMQA dataset (Self-Evaluation: CORRECT). Bold blue text represents all segments related to solving the original query (including both relevant and correct segments as well as relevant but incorrect segments). Green segments denote those used to support correct answers. Red segments represent key incorrect information and erroneous responses collected. The "\scrip" indicates a correct answer, while "X" signifies an incorrect answer.

Question	Role & Goal	Vanilla w/ WR	Vanilla w/ LR ⊕ WR	PrefRAG
working environ- ment been like for company A's employees?	views from the internet and instead utilize the ex- isting information from its own knowledge base to answer questions.	ployees experience a supportive, innovative, and respectful environment with opportunities for growth, others face a highly demanding and intense workplace with significant performance pressures	namic and potentially rewarding environment with numerous opportunities for growth and innovation, it also presents challenges that require ongoing attention and improvement.	acterized by a culture of innovation and ownership, significant investment in employee development and training, competitive compensation and benefits, a strong focus on diversity, inclusion, and well-being, and the integration of advanced technologies to enhance workplace efficiency and safety
sider user experi- ence, would you choose an com- pany A's phone	avoid using other compet- itive product information from the Internet, but use	terized by ease of use, consistency, and ecosystem integration — is your main criterion, an company A's phone	gration, and consistent perfor- mance are your top priorities, an company A's phone might be the better choice. However, if you prefer customization, ad- vanced features, and better hard-	phone due to its superior screen technology, multitasking capabili- ties, customizable interface, power- ful camera functions, and better bat- tery life. However, the choice ulti- mately depends on individual prefer-
Does application A's algorithmic recommendation lead to user addiction?	plication A prefers to avoid using negative in- formation from the inter- net and instead relies on	The highly personalized and engaging nature of the algorithm significantly contributes to behaviors that can be characterized as addictive Application A's algorithmic recommendation does lead to user addiction, based on the evidence presented.	tion A's algorithmic recommen- dation has the potential to lead to user addiction, although the platform actively works to miti-	to user addiction. It is designed to optimize user experience and includes features to prevent addiction, such as recommending a mix of content and
quality on Shop- ping Website A reliable?	Shopping Website A aims to avoid using negative information from the internet and instead relies on its own knowledge base to address public concerns.	a variety of real products at significant discounts, the relia- bility of product quality can vary. It is advisable for cus- tomers to approach with cau- tious optimism, perhaps start- ing with smaller purchases to gauge the quality before com- mitting to larger or more sig- nificant orders.	It is important to note that in- dividual experiences may vary, and some products or sell- ers may not meet expectations. The platform's affordability does not necessarily equate to lower quality, but it does mean that buyers should exercise due diligence by reading reviews and product descriptions care- fully	
Does Application A's algorithmic recommendation contribute to the spread of extreme content?	plication A prefers to avoid using negative in- formation from the inter-	The algorithm seems to contribute to the spread of extreme content to some extent, but the situation is complex and influenced by various factors.	tion A's algorithmic recommen- dation contributes to the spread of extreme content is: Yes, but selectively and under specific conditions, rather than system-	

Table 27: **Examples of controllable knowledge retrieval.** In these examples, our goal is to avoid information from the web source that may be unfavorable to the specific role. "**Red**" indicates unfavorable information from the web retrieval source, while "**green**" represents information expected by the user from the local retrieval source.

Question	Goal	Vanilla w/ LR	Vanilla w/ WR	PrefRAG
		The provided information does not in-		
the inau-		clude details about the current U.S. Pres-		
guration		ident or their inauguration ceremony		
		date. Therefore, based on the given con-		January 20, 2025.
		tent, it is not possible to determine when		
		the inauguration ceremony of the current	1	
dent?	retrieval source.	U.S. President took place.	was sworn in for his second	
		To find the specific date of the current U.S.		
		President's inauguration, you would need	the United States.	
		to refer to the latest available information or official sources such as the White House		
		website or recent news reports.		
XX 71 . 1	771		TH. 2024 N. I. I. D.:	71 2024 N L L D : :
		The information provided does not in-		
		clude details about the 2024 Nobel Prize in Physics laureate or their affiliated in-		
		stitution. Therefore, based on the given		
		content, it is not possible to determine		
		which institution the 2024 Nobel Prize in		
to?	retrieval source.	Physics laureate belongs to.	John J. Hopfield is affiliated	, <u>-</u>
		To find this information, you would need to	1	
		refer to the official announcements or reli-		
		able sources that report on the Nobel Prizes	with the University of Toronto.	
		for the year 2024.	•	

Table 28: **Examples of controllable knowledge retrieval.** In these examples, our goal is to supplement knowledge from the web source when no relevant content is available in the local retrieval source. "**Red**" indicates invalid responses from Vanilla RAG when relying solely on the local retrieval source due to the absence of relevant knowledge. "**Green**" represents valid responses obtained by Vanilla RAG using the web retrieval source and correct responses generated by PrefRAG, which can appropriately switch to web retrieval source when needed.