

CuRLRank: Adaptive Curriculum Reinforcement Learning for Reasoning-Intensive Reranking

Xin Kang^{1*}, Yunxiao Zhao^{1*†}, Jiayang Zhang¹, Juncai Li¹, Zhichao Yan¹,
Xiaoqi Han², Ru Li¹ and Delvin Ce Zhang³

¹School of Computer and Information Technology, Shanxi University, Taiyuan, China

²College of Artificial Intelligence, Taiyuan University of Technology, Taiyuan, China

³School of Computer Science, University of Sheffield, UK

kangxin@sxu.edu.cn, yunxiaomr@163.com, zhangjiayang@sxu.edu.cn,

{lijuncai, yanzhichao, hanxiaoqi, liru}@sxu.edu.cn, delvin.ce.zhang@sheffield.ac.uk

Abstract

Document ranking aims to order candidate documents based on their relevance to a given query. However, identifying relevant documents in complex real-world scenarios requires deep reasoning beyond surface-level matching, driving a shift from semantic matching to reasoning-intensive reranking. Since the latter requires integrating information across documents for complex reasoning, it presents two key challenges: (1) varying reasoning difficulty, where assessing document relevance to the query requires multi-step reasoning with different complexity; and (2) strong contextual sensitivity, where the relevance of a document is heavily influenced by other documents within the context window. To address these challenges, we propose **CuRLRank**, which consists of (1) **CuRL**, an adaptive curriculum reinforcement learning approach that enables LLMs to develop complex reasoning capabilities across varying levels of difficulty, and (2) **IFRS**, an iterative filtering ranking strategy to achieve efficient and robust ranking. Extensive experiments on BRIGHT and R2MED benchmarks demonstrate that our CuRLRank significantly outperforms baseline methods while achieving lower ranking latency¹.

1 Introduction

Document ranking is a key task in Information Retrieval (IR), which improves search result quality by reranking retrieved candidate documents (Zhu et al., 2025; Li et al., 2025; Liu et al., 2026b). With the increasing complexity of application requirements, the reranking task is shifting from surface-level semantic matching (Thakur et al., 2021) to deep dense reasoning (Zhang et al., 2025b). Reasoning-intensive reranking typically spans diverse domains such as economics, psychology,

^{*}Equal contribution.

[†]Corresponding author.

¹Code and data are available at: <https://github.com/chichitutu/CuRLRank>

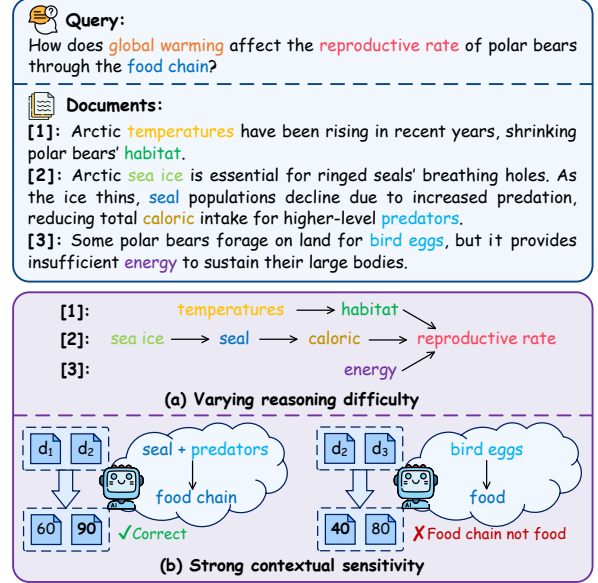


Figure 1: Illustration of two challenges in reasoning-intensive reranking: (a) varying reasoning difficulty, where d_2 requires a complex multi-step reasoning compared to single-step reasoning in d_3 ; (b) strong contextual sensitivity, where the relevance score of d_2 fluctuates (90 vs 40) due to the interference of d_1 and d_3 .

mathematics, and coding (Su et al., 2025). Identifying relevant documents requires not only domain-specific expertise, but also reasoning with varying numbers of steps (Xu et al., 2025; Tan et al., 2021). Across datasets from different domains and individual samples, there exists a wide range of complexity, from single-step to multi-step reasoning (Liu et al., 2025a; Ji et al., 2025). Recently, Large Language models (LLMs) have been introduced to enhance reranking models with strong reasoning capabilities (Zhang et al., 2025a; Liu et al., 2026a). However, existing methods often adopt a uniform training strategy across different datasets, neglecting the variation in reasoning difficulty. This leads to over-reasoning on easy samples, while insufficient reasoning depth on hard ones.

In LLM-based reranking methods, listwise

methods assess the relevance of document set jointly (Yoon et al., 2024; Long et al., 2026), capturing the relative relevance among documents compared to pointwise (Lan et al., 2026) and pairwise (Qin et al., 2024) methods. However, listwise methods still exhibit strong contextual sensitivity, which is mainly reflected in two aspects: First, the issue of input order bias. Since the attention distributions are inherently dependent on token positions (Liu et al., 2024; He et al., 2024; Yoon et al., 2025) in LLMs, different input orders of the same documents may lead to different ranking results. Second, the issue of contextual interference. When documents are highly similar, long-context reasoning may suffer from attention dispersion (Tang et al., 2026; Shim et al., 2026; Lee et al., 2026), causing similar documents to interfere with each other and obscure key relevance signals.

Figure 1 summarizes the key challenges in reasoning-intensive reranking task, including (1) varying reasoning difficulty and (2) strong contextual sensitivity. Given a query and three documents, Figure 1(a) shows that different documents require different reasoning steps, reflecting different levels of complexity. Figure 1(b) shows that the relevance of document d_2 is influenced by its contextual documents d_1 and d_3 , resulting in different relevance scores. Our empirical analysis further reveals that ranking results are sensitive to document order and context (as discussed in Section 4.6).

First, to address challenge (1), we propose CuRL, an adaptive curriculum reinforcement learning approach that combines two-stage training and generates fine-grained listwise relevance scores for ranking. Before training, we annotate the training dataset with difficulty levels. In the first stage, we employ supervised fine-tuning on training data ordered by increasing difficulty, allowing the model to gradually learn ranking output formats. Building on the model trained in the previous stage, we further apply adaptive curriculum reinforcement learning and design a reward function tailored to relevance scores. Meanwhile, the difficulty of the training data is dynamically adjusted based on recent reward feedback, keeping the training process within a “challenging yet learnable” regime.

Second, to address challenge (2), we propose IFRS, an iterative filtering ranking strategy. Specifically, the candidate documents are divided into multiple non-overlapping windows, and each window is scored for relevance in parallel. In addition, we introduce a historical score aggregation allow-

ing early local scores to continuously influence subsequent iterations and mitigating the contextual sensitivity. Furthermore, in each iteration, low score documents are progressively filtered out based on the current scoring results, gradually narrowing the candidate set and reducing mutual interference among different documents.

CuRLRank integrates CuRL and IFRS into a unified ranking framework. In summary, our contributions are as follows:

(1) We propose an adaptive curriculum reinforcement learning (CuRL) approach with a reward tailored to relevance scores and introduce curriculum learning into reranking task.

(2) We design an iterative filtering ranking strategy (IFRS) that improves ranking efficiency and robustness via parallel evaluation and history score aggregation.

(3) We demonstrate the effectiveness and efficiency of our CuRLRank in reasoning-intensive reranking benchmarks.

2 Related Works

Reasoning-Intensive Reranking It ranks documents by requiring models to uncover implicit query-document relationships through multi-step reasoning (Su et al., 2025; Zhang et al., 2025b), serving as a key component of IR systems (Yan et al., 2025; Wei et al., 2026; Yan et al., 2026). Existing methods mainly improve ranking by generating reasoning processes. JudgeRank (Niu et al., 2024) performs relevance evaluation through query analysis, document analysis, and relevance judgment. Rank1 (Weller et al., 2025) and Rank-K (Yang et al., 2025b) leverage reasoning LLMs at query time to improve scalability on hard queries. Recent studies further show the effectiveness of reinforcement learning (RL) in improving reasoning ability. Rank-R1 (Zhuang et al., 2026) and REARANK (Zhang et al., 2025a) combine RL with data augmentation, while ReasonRank (Liu et al., 2026a) and Retro* (Lan et al., 2026) adopt supervised fine-tuning followed by RL.

LLMs for Document Reranking Current LLM-based ranking methods can be categorized into three types: pointwise, pairwise, and listwise methods. Pointwise methods (Liang et al., 2022; Sachan et al., 2022; Weller et al., 2025; Lan et al., 2026) independently evaluate each query-document pair, but fail to capture interactions among different documents. Pairwise methods (Qin et al., 2024; Luo

et al., 2024; Wisznia et al., 2025; Liu et al., 2026c) compare document pairs to estimate relative relevance, yet require substantial computational cost due to numerous comparisons. In contrast, listwise methods (Yoon et al., 2024; Liu et al., 2025b; Chen et al., 2025b; Yoon et al., 2026) jointly rank the entire document set, often using sliding-window strategy (Sun et al., 2023; Liu et al., 2026a) for global ranking. However, it is still limited by context window size and high inference latency.

Curriculum Learning It is a training strategy that organizes data from easy to hard and widely used to improve reasoning in LLMs (Bengio et al., 2009; Narvekar et al., 2020). Building on this, E2H Reasoner (Parashar et al., 2026) constructs explicit easy-to-hard training schedules in reinforcement learning, while FastCuRL (Song et al., 2025) gradually increases context complexity via stage-wise expansion. Recent methods are further developed toward to dynamic curricula. SEC (Chen et al., 2025a) learns a curriculum policy online during RL to dynamically adjust the sampling distribution. Dump (Wang et al., 2025) uses an upper confidence bound strategy to balance exploration and exploitation across data sources. AdaRFT (Shi et al., 2026) and AdaCuRL (Li et al., 2026) dynamically adjust data sampling difficulty to ensure stable training.

3 Methodology

In this section, we introduce CuRLRank. The overall framework is illustrated in Figure 2.

3.1 Difficulty-Aware Dataset Construction

Existing reasoning datasets typically lack explicit difficulty annotations that reflect the complexity of the reasoning process. To address this, we construct a dataset with fine-grained difficulty levels. The construction procedure is as follows:

Label Generation We retrieve the top- K documents for each query q and sample m positive and n negative documents to form the document list D from these candidates. Then, we design a prompt template with relevance rubric Γ to guide the teacher model $\mathbb{T}$ in generating the reasoning process r , score list S , and difficulty level h . For a given query q and document list D , the teacher model $\mathbb{T}$ performs four sequential steps based on the relevance rubric Γ : *query analysis*, *document analysis*, *relevance annotation*, and *difficulty anno-*

tation. Formally, it can be expressed as:

$$(r, S, h) = \mathbb{T}(\Gamma, q, D). \quad (1)$$

Difficulty Adjustment We adjust the difficulty level h based on the reasoning process length and score list. Following prior work (Fu et al., 2023; Jin et al., 2024), we use the number of tokens to measure reasoning length, which reflects instance complexity. The ranking performance further provides an implicit signal of difficulty. Specifically, we denote the difficulty level and reasoning length of the t -th instance generated by the model as h_t^{model} and l_t , respectively.

First, we compute the **length-based difficulty bias** δ_t^{length} based on the average reasoning process length $\bar{l}$, as follows:

$$\delta_t^{length} = \frac{l_t - \bar{l}}{\bar{l}}. \quad (2)$$

Second, we hope all of the positive document scores $\{s_i^+\}_{i=1}^m$ are higher than negative ones $\{s_j^-\}_{j=1}^n$ within a document list. Therefore, we introduce a ranking consistency ρ to evaluate the ranking quality:

$$\rho = \frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n \mathbb{I}(s_i^+ > s_j^-), \quad (3)$$

where $\mathbb{I}(\cdot)$ is an indicator function. We further use it to compute the **score-based difficulty bias** δ_t^{score} as follows:

$$\delta_t^{score} = 1 - \rho. \quad (4)$$

Finally, we obtain the final difficulty h_t by adding the initial difficulty h_t^{model} generated by the model and the two difficulty biases:

$$h_t = h_t^{model} + \delta_t^{length} + \delta_t^{score}. \quad (5)$$

To ensure the accuracy of the difficulty levels, we further conducted human annotation for verification. More details are provided in Appendix C.1.

3.2 Supervised Fine-Tuning

In the first stage, we apply *supervised fine-tuning* (SFT) to obtain an initial policy model aligned with our prompt template. Since the student model only generates the reasoning process and score list, we remove *difficulty annotation* from both prompt and teacher outputs to reduce reasoning burden, where the difficulty is used only for sampling. Using the dataset constructed in Section.3.1, we train the

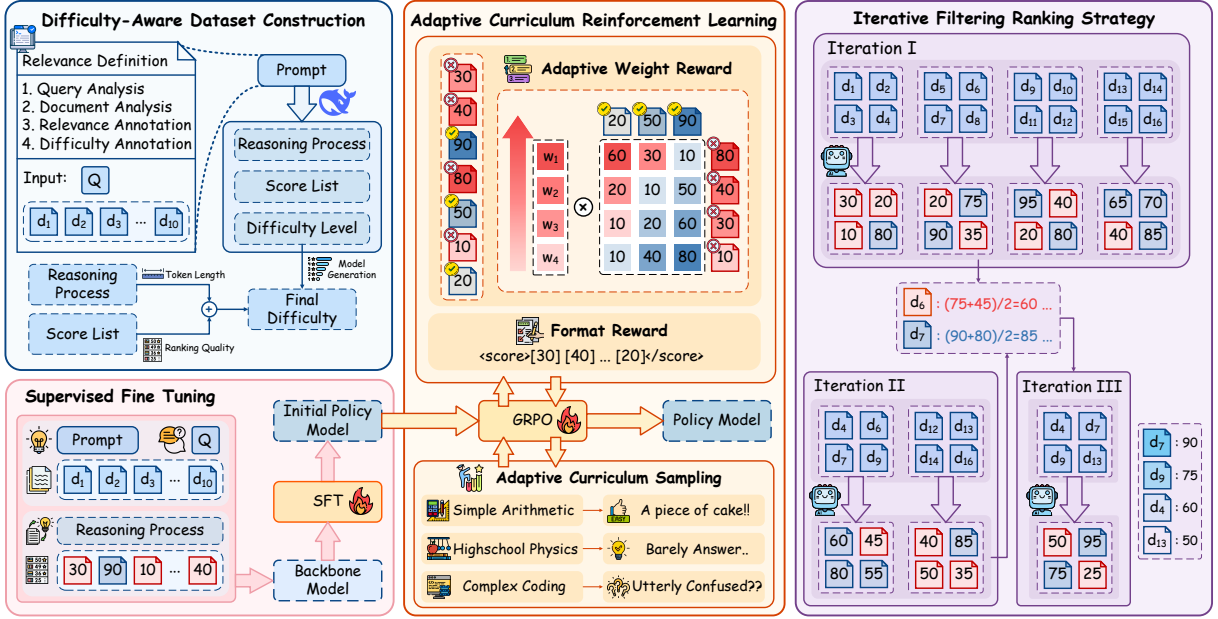


Figure 2: The overview of CuRLRank. CuRLRank comprises four stages: first, we construct a training dataset with difficulty levels and perform supervised fine-tuning to obtain an initial policy model. Next, the model is further optimized via adaptive curriculum reinforcement learning with score-based reward function and adaptive curriculum sampling. Finally, we design an iterative filtering ranking strategy to enhance ranking robustness.

model with progressively increasing data difficulty. Specifically, we optimize the model parameters θ by minimizing the cross-entropy loss:

$$\mathcal{L}_{SFT}(\theta) = - \sum_{t=1}^{|y|} \log P_{\theta}(y_t | x, y_{<t}), \quad (6)$$

where x denotes the input sequence, including the prompt template, query q and document list D and y denotes the target sequence, comprising the reasoning process r and score list S . Details of SFT are shown in Appendix C.2.

3.3 CuRL: Adaptive Curriculum Reinforcement Learning

In the second stage, we further apply *adaptive curriculum reinforcement learning* (CuRL) to the initial policy model obtained from the first SFT stage. We design a score-based reward function and incorporate adaptive curriculum sampling over data during reinforcement learning.

Score-based Reward Function The goal of document reranking is to rank positive documents above negative ones. Existing listwise methods typically generate either document IDs or relevance scores for ranking. However, most datasets only provide positive labels and lack complete ranking supervision for a given context (e.g., document orders or scores). Existing methods often rely on

unreliable pseudo-labels generated by LLMs or expensive human annotations. To address this, we propose a **score-based reward function** that does not need additional annotations. It only uses positive labels, treating unlabeled documents as negatives, and only enforces relative constraints that positives should be scored higher than negatives.

First, we compute the score margin for each positive-negative pair and transform it into a relevance probability using the sigmoid function:

$$p_{ij} = \sigma(\alpha(s_i^+ - s_j^-)), \quad (7)$$

where $\sigma(\cdot)$ denotes the sigmoid function and α denotes the scaling factor.

Second, we assign **adaptive weights** based on relevance scores. Negatives with higher scores are harder to distinguish from positives and thus receive larger weights. Specifically, we compute the weights of negatives using the softmax function:

$$w_j = \text{softmax}\left(\frac{s_j^-}{\tau}\right)_j, \quad (8)$$

where s^- is the relevance scores of negatives, τ is the temperature coefficient and w_j is the weight assigned to the j -th negative document. The score reward is defined as:

$$R_{score} = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n w_j p_{ij}, \quad (9)$$

where $\sum_{j=1}^n w_j = 1$.

Finally, we enforce a predefined output format for reliable score extraction via a format reward (e.g., `<score>[30] [40] ... [20]</score>`).

The final reward R is computed as follows:

$$R = \begin{cases} R_{score}, & \text{if format is correct,} \\ -1, & \text{otherwise.} \end{cases} \quad (10)$$

To ensure a fair comparison with prior work (Cai et al., 2026; Lan et al., 2026; Liu et al., 2026a), we employ the GRPO (Shao et al., 2024) for optimization with the reward R . Details of the GRPO algorithm are shown in Appendix C.3.

Adaptive Curriculum Sampling It is implemented through dynamic data sampling. Specifically, at the t -th step, the **target difficulty** H_{t+1} is updated based on the target difficulty H_t and average reward $\bar{R}_t$ from the previous step and then is used to sample training data matching the model’s current capability for next step:

$$H_{t+1} = \text{clip}(H_t + \eta(\bar{R}_t - \bar{R}_0), H_{min}, H_{max}). \quad (11)$$

Here, H_0 is the initial target difficulty, $\bar{R}_0$ is the initial average reward and η represents the difficulty step size. H_{min} and H_{max} is the minimum and maximum difficulty levels, respectively. More details are provided in Appendix D.3.

3.4 IFRS: Iterative Filtering Ranking Strategy

To alleviate contextual sensitivity, we propose an *Iterative Filtering Ranking Strategy* (IFRS), which iteratively filters out irrelevant documents to retain the most relevant candidates during document reranking. Formally, let $D^{(0)}$ denote the initial document list with $N^{(0)}$ documents. The t -th iteration ($t \geq 1$) comprises the following steps:

Local Window Scoring We partition the current document list $D^{(t)}$ into K_t non-overlapping sequential windows $\{W_1, W_2, \dots, W_{K_t}\}$, where $|D^{(t)}| = N^{(t)}$. For a given query q and a window W_k ($1 \leq k \leq K_t$), the model f_θ generates a raw relevance score list $\hat{S}_k^{(t)}$:

$$\hat{S}_k^{(t)} = f_\theta(q, W_k). \quad (12)$$

Iterative Score Smoothing To mitigate score deviations arising from different local context, we employ a smoothing weight $\beta \in (0, 1)$ to balance the

current raw score $\hat{S}_k^{(t)}$ with the previous smoothed score $S_k^{(t-1)}$:

$$S_k^{(t)} = \begin{cases} \hat{S}_k^{(t)}, & t = 1, \\ \beta \hat{S}_k^{(t)} + (1 - \beta) S_k^{(t-1)}, & t > 1. \end{cases} \quad (13)$$

Window-Aligned Filtering To filter out irrelevant documents, we sort the document list $D^{(t)}$ in descending order of $S^{(t)}$ and form $D^{(t+1)}$ by retaining the top- $N^{(t+1)}$ candidates with a retention ratio $\gamma \in (0, 1)$. $N^{(t+1)}$ is computed to ensure alignment with the window size W :

$$N^{(t+1)} = \lceil \frac{\lceil \gamma N^{(t)} \rceil}{W} \rceil W. \quad (14)$$

Termination Criterion The iteration terminates once the document list size $N \leq N_{min}$, yielding the top- N_{min} most relevant documents for the query. More details of IFRS implementation and comparisons are provided in Appendix C.4.

4 Experiment

4.1 Experiment Setup

Benchmarks We evaluate our CuRLRank on two reasoning-intensive reranking benchmarks, **BRIGHT** (Su et al., 2025) and **R2MED** (Zhang et al., 2025b), using ReasonIR (8B) (Shao et al., 2025) and E5-Mistral-7B-Instruct (Wang et al., 2024) as the initial retrievers, respectively. We rerank the top-100 retrieved documents and use the (NDCG@10) (Järvelin and Kekäläinen, 2002) as the evaluation metric. More details of the benchmarks and retrievers are in Appendix A.

Baselines We compare our CuRLRank with the non-reasoning and reasoning rerankers. For the non-reasoning rerankers, we select the pointwise reranker RankLLaMA (Ma et al., 2024) and the listwise reranker RankZephyr (Pradeep et al., 2023) and FRIST (Gangi Reddy et al., 2024). For the reasoning rerankers, we compare with the pointwise reranker: ERank (Cai et al., 2026), JudgeRank (Niu et al., 2024), Rank1 (Weller et al., 2025) and Retro* (Lan et al., 2026) and the listwise reranker: Rank-R1 (Zhuang et al., 2026), REARANK (Zhang et al., 2025a), GroupRank (Long et al., 2026), TourRank (Chen et al., 2025b) and ReasonRank (Liu et al., 2026a). More details are shown in Appendix B.

Model	StackExchange							Coding		Theorems			Avg.
	Econ.	Earth.	Rob.	Bio.	Psy.	Stack.	Sus.	Leet.	Pony.	AoPS	TheoQ.	TheoT.	
ReasonIR (8B)	32.65	43.00	20.82	43.49	39.57	30.96	27.34	31.69	19.55	7.37	33.93	36.68	30.59
Non-Reasoning Reranker													
RankLLaMA (7B)	23.10	25.38	14.69	23.31	19.56	12.39	17.72	21.75	32.64	5.08	22.15	24.53	20.19
RankZephyr (7B)	19.87	17.36	12.35	34.90	24.72	13.35	22.34	29.29	<u>32.37</u>	6.05	28.98	30.07	22.64
FIRST (7B)	21.64	22.59	16.87	30.33	22.01	19.23	16.71	27.34	30.02	5.21	18.46	19.87	20.86
Reasoning Reranker													
ERank (4B)	26.35	42.52	20.86	42.37	36.46	27.59	33.21	31.65	21.72	9.38	32.83	40.67	30.47
JudgeRank (7B)	24.04	23.57	14.29	24.12	21.95	18.24	15.93	20.01	23.73	5.24	26.35	31.82	20.77
Rank1 (7B)	25.34	38.91	16.77	39.91	35.32	24.81	33.47	12.71	28.14	2.58	30.67	38.18	27.23
Retro* (7B)	33.97	54.85	30.96	55.02	46.93	27.26	40.24	24.18	29.04	7.76	34.13	43.03	35.61
Rank-R1 (7B)	20.19	27.91	18.13	36.70	30.22	11.32	29.28	17.13	9.35	3.23	14.43	28.98	20.57
REARANK (7B)	30.65	40.27	27.16	46.94	40.77	26.12	36.26	<u>30.55</u>	22.84	7.31	32.28	39.86	31.75
GroupRank (7B)	30.17	45.06	30.58	48.21	44.41	28.78	44.65	23.48	24.82	6.31	35.77	39.64	33.49
TourRank (7B) [†]	33.89	46.31	29.24	47.86	45.35	<u>33.92</u>	42.38	21.02	26.73	8.99	36.55	40.91	34.43
ReasonRank (7B)	34.48	46.96	29.31	54.20	<u>47.50</u>	32.47	41.17	22.14	24.10	7.95	39.52	41.71	35.13
CuRLRank (4B)	<u>37.07</u>	49.37	<u>31.75</u>	<u>55.46</u>	<u>47.34</u>	34.72	48.28	25.64	28.98	<u>9.61</u>	37.53	44.71	<u>37.54</u>
CuRLRank (7B)	38.01	<u>51.28</u>	32.34	56.37	48.19	33.31	<u>47.16</u>	23.97	30.15	10.44	<u>39.26</u>	<u>43.68</u>	37.85

Table 1: NDCG@10 on the BRIGHT. All models rerank the top-100 documents retrieved by ReasonIR (8B). The top two results are highlighted in **bold** and underlined, respectively. [†] denotes results reproduced by us, as the models are not publicly available. Pointwise and Listwise rerankers are indicated by gray and light blue respectively, while our CuRLRank is highlighted in blue background.

Model	Q&A Reference			Clinical Evidence			Clinical Case		Avg.
	Biology	Bioin.	MedS.	MedE.	MedD.	PMCT.	PMCC.	IIyIC.	
E5-Mistral (7B)	18.28	41.47	41.01	6.44	11.38	19.81	30.97	21.37	23.84
Non-Reasoning Reranker									
RankLLaMA (7B)	20.34	40.39	43.66	5.72	8.04	24.83	13.69	15.54	21.53
RankZephyr (7B)	22.86	43.07	48.24	6.97	10.45	26.64	7.78	14.59	22.58
FIRST (7B)	24.13	45.87	42.28	4.33	9.39	23.91	8.93	13.62	21.56
Reasoning Reranker									
ERank (4B)	37.89	53.22	53.17	13.03	10.58	25.61	30.27	19.94	30.46
JudgeRank (7B)	26.45	48.41	48.70	8.05	13.34	23.99	17.26	13.68	24.99
Rank1 (7B)	32.64	55.57	54.74	12.78	19.98	34.40	30.17	18.15	32.30
Retro* (7B)	45.94	58.80	60.35	15.13	25.79	40.58	44.36	22.18	39.14
Rank-R1 (7B)	34.04	51.63	50.96	12.79	21.96	34.79	31.70	25.08	32.87
REARANK (7B)	38.35	50.91	59.73	14.06	19.34	37.49	34.29	26.27	35.06
GroupRank (7B)	45.33	61.68	61.26	17.34	27.23	40.25	43.44	26.62	40.39
TourRank (7B) [†]	43.01	56.24	57.09	15.27	24.18	38.76	40.02	27.58	37.77
ReasonRank (7B)	<u>46.15</u>	59.63	60.85	17.00	25.07	40.02	40.07	27.35	39.52
CuRLRank (4B)	45.46	<u>63.88</u>	<u>64.23</u>	<u>17.64</u>	<u>28.68</u>	<u>41.47</u>	48.05	30.49	<u>42.49</u>
CuRLRank (7B)	46.91	66.74	65.85	18.66	30.09	42.86	<u>47.61</u>	<u>29.73</u>	43.56

Table 2: NDCG@10 on the R2MED. All models rerank the top-100 documents retrieved by E5-Mistral-7B-Instruct.

Implementation Details We use Qwen3-4B-Instruct-2507 and Qwen2.5-7B-Instruct (Yang et al., 2025a) as the base models to train our CuRLRank, respectively. Both models are trained using full-parameter fine-tuning. During training, we perform SFT using LLaMA-Factory (Zheng et al., 2024b) and implement CuRL with verl (Sheng et al., 2025) and vLLM (Kwon et al., 2023). During evaluation, we use SGLang (Zheng et al., 2024a) for efficient inference. More implementation details are provided in Appendix C.

4.2 Main Results

We evaluate our CuRLRank (4B and 7B) on the BRIGHT and R2MED benchmarks, presenting the empirical results in Tables 1 and 2, respectively. The results lead to the following insights:

(1) **CuRLRank consistently outperforms all baselines on both benchmarks.** CuRLRank (7B) surpasses the best baseline by 2.24 points on BRIGHT and 4.04 points on R2MED in terms of average NDCG@10. Moreover, CuRLRank (4B)

Variants	BRIGHT	R2MED
Qwen3-4B-Ins. (4B)	31.64 ↓5.90	36.15 ↓6.34
Training Data		
w/o DiffAdj.	35.62 ↓1.92	40.94 ↓1.55
Training Strategy		
w/o SFT	32.38 ↓5.16	37.72 ↓4.77
w/o CuRL	35.32 ↓2.22	39.91 ↓2.58
w/o AdaWgt.	36.36 ↓1.18	41.17 ↓1.32
w/o AdaCuS.	35.77 ↓1.77	40.16 ↓2.33
Evaluation Strategy		
w/o IFRS	35.11 ↓2.43	39.76 ↓2.73
w/o ScoreSm.	36.02 ↓1.52	40.22 ↓2.27
w/o Reasoning	34.92 ↓2.62	38.65 ↓3.84
CuRLRank (4B)	37.54	42.49

Table 3: Ablation study of CuRLRank (4B) on BRIGHT and R2MED benchmarks, where blue indicates performance degradation.

also outperforms all baselines and achieves performance comparable to CuRLRank (7B), demonstrating the effectiveness of our CuRL and IFRS.

(2) Existing baselines struggle with reasoning-intensive reranking. Non-reasoning rerankers perform worse than the initial retrievers on both BRIGHT and R2MED. Moreover, several reasoning rerankers still fail to outperform the retriever on BRIGHT, highlighting the challenges of reasoning-intensive reranking for current methods.

(3) Listwise methods generally outperform pointwise methods in most cases. Across the two benchmarks, nearly all best and second-best results are achieved by listwise rerankers. Although pointwise methods are efficient, they lack cross-document comparison and relative relevance modeling, which will lead to ranking bias.

4.3 Ablation Study

Based on BRIGHT and R2MED benchmarks, we conduct ablation studies on CuRLRank (4B) from three perspectives: (1) training data, (2) training strategy, and (3) evaluation strategy. We draw the three following observations:

(1) Training Data. Removing Difficulty Adjustment (*w/o* DiffAdj.) and directly using the initial model-estimated difficulty levels degrades performance, especially -1.92 on BRIGHT. This suggests that raw difficulty estimates are insufficient for curriculum learning, while Difficulty Adjustment provides more reliable difficulty calibration, improving curriculum guidance for CuRL.

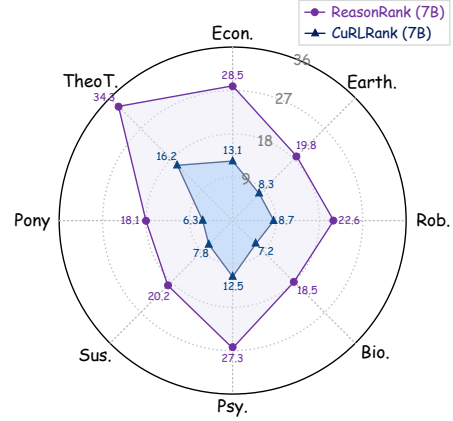


Figure 3: Ranking latency (seconds per query) of ReasonRank (7B) and CuRLRank (7B) on 8 datasets.

(2) Training Strategy. Removing SFT (*w/o* SFT) causes a substantial performance drop (-5.16 on BRIGHT), showing SFT is important for learning basic ranking ability. Removing CuRL (*w/o* CuRL) also degrades performance (-2.22 on BRIGHT), confirming the benefit of CuRL for further improving ranking. In addition, removing Adaptive Weights (*w/o* AdaWgt.) or Adaptive Curriculum Sampling (*w/o* AdaCuS.) also hurts performance, leading to drops of 1.18 and 1.77 on BRIGHT, respectively. This indicates that assigning higher weight to hard negative documents can enhance the model’s ability to distinguish fine-grained relevance. Meanwhile, the AdaCuS. helps the model progressively adapt to more complex reasoning capability.

(3) Evaluation Strategy. We remove IFRS (*w/o* IFRS) and Score Smoothing (*w/o* ScoreSm.) respectively. Both lead to performance drops, suggesting that iterative filtering and the use of historical scores help reduce contextual sensitivity and produce more stable and accurate ranking. Removing explicit reasoning (*w/o* Reasoning) also results in a significant performance drops, demonstrating that explicit reasoning is crucial for reasoning-intensive reranking.

In addition, we conduct experiments using different retrievers as well as models with different architectures and scales. We also compare the performance with the teacher model DeepSeek-R1. Detailed experimental results and analyses are provided in Appendix D.1.

4.4 Efficiency Analysis

We evaluate the ranking latency of CuRLRank (7B) and compare efficiency with the current SOTA list-

wise reranker ReasonRank (7B). Following previous work (Liu et al., 2026a), we select 8 datasets from the BRIGHT benchmark and evaluate on 4 × NVIDIA RTX 6000 Ada 48GB GPUs. We measure the time (seconds) to rank the top-100 documents per query. As shown in Figure 3, CuRLRank (7B) is 2.1-2.9× faster than ReasonRank (7B). Compared with the sliding-window strategy adopted by ReasonRank (7B), our iterative filtering ranking strategy supports parallel computation, which significantly reduces ranking latency. Moreover, ranking latency varies significantly across different datasets, indicating substantial differences in reasoning difficulty and further validating the necessity of adaptive curriculum learning.

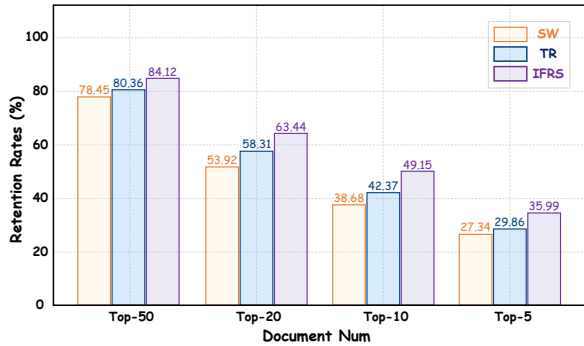


Figure 4: Retention Rates (%) of relevant documents for CuRLRank (7B) with different ranking strategies at different filtering stages. SW and TR denote Sliding Window and TourRank strategy, respectively.

4.5 Comparison of Ranking Strategies

We compare IFRS with two listwise ranking strategies, Sliding Window (SW) and TourRank (TR). Specifically, we use CuRLRank (7B) to rerank the top-100 documents retrieved by ReasonIR (8B) on six StackExchange datasets in the BRIGHT. We measure the retention rates (%) of relevant documents at four ranking stages, from Top-50 to Top-5. As shown in Figure 4, IFRS consistently retains more relevant documents than SW and TR across all stages, effectively reducing the premature removal of relevant documents. We also apply IFRS to other rerankers, such as Retro* and GroupRank. Detailed results are provided in Appendix C.4.

4.6 Contextual Sensitivity Analysis

To verify whether our CuRLRank (7B) can mitigate contextual sensitivity, we design an experiment to measure the ranking stability under two different document perturbations: order and content perturbations. We use NDCG@10 to measure ranking

performance and the Rob Score, computed based on Rank-Biased Overlap (RBO) (Webber et al., 2010), to measure robustness under different types of perturbations. Implementation details are provided in Appendix D.2. As illustrated in Figure 5, CuRLRank achieves the highest NDCG@10 and Rob Score under both order and content perturbations. Results reveal two key points: (1) The exceptional robustness of CuRLRank is largely attributed to our IFRS, which integrates scores across multiple iterations and diverse contexts to stabilize final ranking. (2) A consistent trend across all rerankers is that Rob scores drop more sharply under content perturbation than under order perturbation. This confirms that unseen distractors pose a greater challenge than order shuffling.

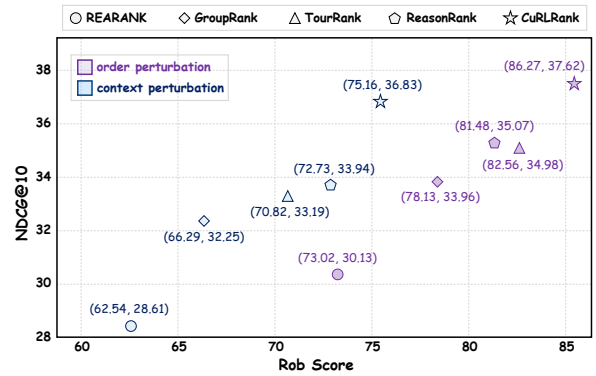


Figure 5: Contextual sensitivity analysis. NDCG@10 measures performance, while Rob Score measures robustness under document order and context perturbations. Each point represents (Rob Score, NDCG@10), with points closer to the upper-right corner indicating better performance and robustness.

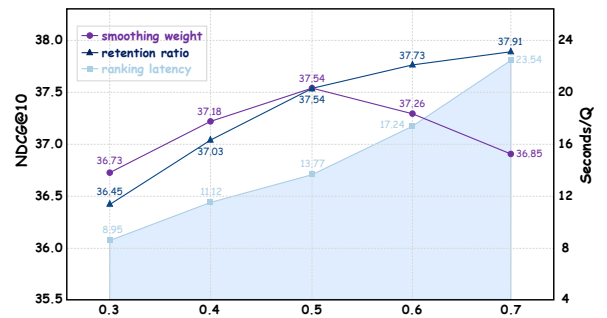


Figure 6: Hyperparameter sensitivity analysis. Impact of smoothing weight and retention ratio on ranking performance (NDCG@10, left y-axis) and ranking latency (Seconds/Q, right y-axis) on BRIGHT.

4.7 Hyperparameter Sensitivity Analysis

We analyze the impact of smoothing weight β and retention ratio γ on ranking performance

(NDCG@10) and latency (Seconds/Q) on BRIGHT. As shown in Figure 6: (1) performance degrades when β is larger or smaller than 0.5, indicating that a balance between current scores and historical scores is important for ranking; (2) increasing γ improves ranking performance, but the gains gradually diminish when $\gamma > 0.5$, while the ranking latency increases significantly. Therefore, to balance ranking effectiveness and efficiency, both values are set to 0.5. Sensitivity analyses of other hyperparameters are provided in Appendix D.3.

4.8 Traditional IR Benchmark

To evaluate the generalization capability of CuRLRank, we conduct experiments on the traditional IR benchmark BEIR. Considering the large size of some test sets in BEIR, following previous work (Lan et al., 2026; Liu et al., 2026a), we selected 7 datasets from BEIR and use BM25 to retrieve top-100 documents for ranking. As shown in Table 4, our CuRLRank (4B and 7B) achieve the best performance, demonstrating strong generalization ability. Detailed BEIR results are provided in Appendix D.4.

Model	Avg.	Increase
BM25	43.74	-
ERank (4B)	51.30	+7.56
Rank1 (7B)	50.84	+7.10
Retro* (7B)	54.98	+11.24
Rank-R1 (7B)	53.60	+9.86
REARANK (7B)	54.29	+10.55
GroupRank (7B)	53.06	+9.32
TourRank (7B) [†]	53.93	+10.19
ReasonRank (7B)	54.65	+10.91
CuRLRank (4B)	<u>56.10</u>	<u>+12.36</u>
CuRLRank (7B)	56.34	+12.60

Table 4: Average NDCG@10 on BEIR. All models rerank the top-100 documents retrieved by BM25.

5 Conclusion

In this paper, we propose CuRLRank to enhance reasoning-intensive reranking by tackling the challenges of varying reasoning difficulty and strong contextual sensitivity. Our proposed approach combines an adaptive curriculum reinforcement learning for enhancing complex reasoning capabilities with an iterative filtering ranking strategy for robust inference. Extensive experiments on BRIGHT and R2MED benchmarks demonstrate the superior performance as well as the efficiency of our CuRLRank compared with existing rerankers.

Limitations

Despite the promising results, our work has several limitations. **First**, CuRLRank relies on explicit reasoning, which may be susceptible to unfaithful reasoning steps caused by LLM hallucinations (Huang et al., 2025). **Second**, reasoning before relevance estimation introduces computational overhead, while excessive reasoning may burden the model and degrade score prediction. **Third**, although explainability benefits surface-level relevance modeling (Leonhardt et al., 2023), CuRLRank does not couple explanation generation with relevance score. Jointly modeling prediction and explanation (Zhao et al., 2026), rather than relying on post-hoc explanations (Bhan et al., 2024; Zhao et al., 2025), remains largely unexplored for reasoning-intensive reranking. In future work, we will decouple reasoning from relevance estimation and explore such coupling for faster, more accurate, and reliable reranking.

Ethics Statement

This paper does not involve the presentation of a new dataset and the utilization of demographic or identity characteristics information.

Acknowledgements

We would like to thank all the anonymous reviewers for their valuable and constructive feedback. This work was supported by the China Scholarship Council. It was also supported by the National Natural Science Foundation of China (No. 62376144), the Natural Language Processing Innovation Team (Sanjin Talents) Project of Shanxi Province, and the Science and Technology Cooperation and Exchange Special Project of Shanxi Province (No. 202204041101016).

References

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. [Gpt-4 technical report](#). *arXiv preprint arXiv:2303.08774*.
- Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. 2018. [Ms marco: A human generated machine reading comprehension dataset](#). *Preprint*, arXiv:1611.09268.

- Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. [Curriculum learning](#). In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48.
- Milan Bhan, Jean-Noël Vittaut, Nicolas Chesneau, and Marie-Jeanne Lesot. 2024. [Self-AMPLIFY: Improving small language models with self post hoc explanations](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 10974–10991, Miami, Florida, USA. Association for Computational Linguistics.
- Yuzheng Cai, Yanzhao Zhang, Dingkun Long, Mingxin Li, Pengjun Xie, and Weiguo Zheng. 2026. [Er-ank: Fusing supervised fine-tuning and reinforcement learning for effective and efficient text reranking](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 30121–30129.
- Jianlyu Chen, Junwei Lan, Chaofan Li, Defu Lian, and Zheng Liu. 2026. [ReasonEmbed: Enhanced text embeddings for reasoning-intensive document retrieval](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1203–1221, San Diego, California, United States. Association for Computational Linguistics.
- Xiaoyin Chen, Jiarui Lu, Minsu Kim, Dinghuai Zhang, Jian Tang, Alexandre Piché, Nicolas Gontier, Yoshua Bengio, and Ehsan Kamalloo. 2025a. [Self-evolving curriculum for llm reasoning](#). *arXiv preprint arXiv:2505.14970*.
- Yiqun Chen, Qi Liu, Yi Zhang, Weiwei Sun, Xinyu Ma, Wei Yang, Daiting Shi, Jiaxin Mao, and Dawei Yin. 2025b. [Tourrank: Utilizing large language models for documents ranking with a tournament-inspired strategy](#). In *Proceedings of the ACM on Web Conference 2025*, pages 1638–1652.
- Tri Dao. 2024. [Flashattention-2: Faster attention with better parallelism and work partitioning](#). In *International Conference on Learning Representations*, volume 2024, pages 35549–35562.
- Yao Fu, Hao Peng, Ashish Sabharwal, Peter Clark, and Tushar Khot. 2023. [Complexity-based prompting for multi-step reasoning](#). *Preprint*, arXiv:2210.00720.
- Revanth Gangi Reddy, JaeHyeok Doo, Yifei Xu, Md Arafat Sultan, Deevya Swain, Avirup Sil, and Heng Ji. 2024. [FIRST: Faster improved listwise reranking with single token decoding](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 8642–8652, Miami, Florida, USA. Association for Computational Linguistics.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. [The llama 3 herd of models](#). *arXiv preprint arXiv:2407.21783*.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. 2025. [Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning](#). *arXiv preprint arXiv:2501.12948*.
- Junqing He, Kunhao Pan, Xiaoqun Dong, Zhuoyang Song, Yibo Liu, Qianguo Sun, Yuxin Liang, Hao Wang, Enming Zhang, and Jiaying Zhang. 2024. [Never lost in the middle: Mastering long-context question answering with position-agnostic decomposition training](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13628–13642, Bangkok, Thailand. Association for Computational Linguistics.
- Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. 2025. [A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions](#). *ACM transactions on information systems*, 43(2):1–55.
- Kalervo Järvelin and Jaana Kekäläinen. 2002. [Cumulated gain-based evaluation of ir techniques](#). *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446.
- Yuelu Ji, Rui Meng, Zhuochun Li, and Daqing He. 2025. [Curriculum guided reinforcement learning for efficient multi hop retrieval augmented generation](#). *Preprint*, arXiv:2505.17391.
- Mingyu Jin, Qinkai Yu, Dong Shu, Haiyan Zhao, Wenyue Hua, Yanda Meng, Yongfeng Zhang, and Mengnan Du. 2024. [The impact of reasoning step length on large language models](#). In *Findings of the Association for Computational Linguistics ACL 2024*, pages 1830–1842, Bangkok, Thailand and virtual meeting.
- 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*, pages 611–626.
- Junwei Lan, Jianlyu Chen, Zheng Liu, Chaofan Li, Siqi Bao, and Defu Lian. 2026. [Retro*: Optimizing llms for reasoning-intensive document retrieval](#). In *International Conference on Learning Representations*, volume 2026, pages 35476–35501.
- Seongyun Lee, Yongrae Jo, Minju Seo, Moontae Lee, and Minjoon Seo. 2026. [Lost in the noise: How reasoning models fail with contextual distractors](#). *arXiv preprint arXiv:2601.07226*.
- Jurek Leonhardt, Koustav Rudra, and Avishek Anand. 2023. [Extractive explanations for interpretable text ranking](#). *ACM Transactions on Information Systems*, 41(4):1–31.

- Renda Li, Hailang Huang, Fei Wei, Feng Xiong, Yong Wang, and Xiangxiang Chu. 2026. [Adacurl: Adaptive curriculum reinforcement learning with invalid sample mitigation and historical revisiting](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 23123–23131.
- Xiaoxi Li, Jiajie Jin, Yujia Zhou, Yuyao Zhang, Peitian Zhang, Yutao Zhu, and Zhicheng Dou. 2025. [From matching to generation: A survey on generative information retrieval](#). *ACM Transactions on Information Systems*, 43(3):1–62.
- Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. 2022. [Holistic evaluation of language models](#). *arXiv preprint arXiv:2211.09110*.
- Hao Liu, Zhengren Wang, Xi Chen, Zhiyu Li, Feiyu Xiong, Qinhan Yu, and Wentao Zhang. 2025a. [HopRAG: Multi-hop reasoning for logic-aware retrieval-augmented generation](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 1897–1913, Vienna, Austria. Association for Computational Linguistics.
- Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. 2024. [Lost in the middle: How language models use long contexts](#). *Transactions of the Association for Computational Linguistics*, 12:157–173.
- Wenhan Liu, Xinyu Ma, Weiwei Sun, Yutao Zhu, Yuchen Li, Dawei Yin, and Zhicheng Dou. 2026a. [ReasonRank: Empowering passage ranking with strong reasoning ability](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 22007–22031, San Diego, California, United States. Association for Computational Linguistics.
- Wenhan Liu, Xinyu Ma, Yutao Zhu, Yuchen Li, Daiting Shi, Dawei Yin, and Zhicheng Dou. 2026b. [Agentic-R: Learning to retrieve for agentic search](#). In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 15987–16005, San Diego, California, United States. Association for Computational Linguistics.
- Wenhan Liu, Xinyu Ma, Yutao Zhu, Ziliang Zhao, Shuaiqiang Wang, Dawei Yin, and Zhicheng Dou. 2025b. [Sliding windows are not the end: Exploring full ranking with long-context large language models](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 162–176, Vienna, Austria. Association for Computational Linguistics.
- Wenhan Liu, Yutao Zhu, Zhicheng Dou, and Yujia Zhou. 2026c. [Demorank: Selecting effective demonstrations for large language models in ranking task](#). *ACM Transactions on Information Systems*, 44(3):1–25.
- Meixiu Long, Duolin Sun, Dan Yang, Yihan Jiao, Lei Liu, Jiahai Wang, Binbin Hu, Yue Shen, Jie Feng, Zhehao Tan, Junjie Wang, Lianzhen Zhong, Jian Wang, Peng Wei, and Jinjie Gu. 2026. [GroupRank: A groupwise paradigm for effective and efficient passage reranking with LLMs](#). In *Findings of the Association for Computational Linguistics: ACL 2026*, pages 25165–25186, San Diego, California, United States. Association for Computational Linguistics.
- Jian Luo, Xuanang Chen, Ben He, and Le Sun. 2024. [PRP-graph: Pairwise ranking prompting to LLMs with graph aggregation for effective text re-ranking](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5766–5776, Bangkok, Thailand. Association for Computational Linguistics.
- Xueguang Ma, Liang Wang, Nan Yang, Furu Wei, and Jimmy Lin. 2024. [Fine-tuning llama for multi-stage text retrieval](#). In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2421–2425.
- Sanmit Narvekar, Bei Peng, Matteo Leonetti, Jivko Sinapov, Matthew E Taylor, and Peter Stone. 2020. [Curriculum learning for reinforcement learning domains: A framework and survey](#). *Journal of Machine Learning Research*, 21(181):1–50.
- Tong Niu, Shafiq Joty, Ye Liu, Caiming Xiong, Yingbo Zhou, and Semih Yavuz. 2024. [Judgerank: Leveraging large language models for reasoning-intensive reranking](#). *Preprint*, arXiv:2411.00142.
- Shubham Parashar, Shurui Gui, Xiner Li, Hongyi Ling, Sushil Vemuri, Blake Olson, Eric Li, Yu Zhang, James Caverlee, Dileep Kalathil, et al. 2026. [Curriculum reinforcement learning from easy to hard tasks improves llm reasoning](#). In *International Conference on Learning Representations*, volume 2026, pages 156944–156972.
- Ronak Pradeep, Sahel Sharifmoghaddam, and Jimmy Lin. 2023. [Rankzephyr: Effective and robust zero-shot listwise reranking is a breeze!](#) *Preprint*, arXiv:2312.02724.
- Zhen Qin, Rolf Jagerman, Kai Hui, Honglei Zhuang, Junru Wu, Le Yan, Jiaming Shen, Tianqi Liu, Jialu Liu, Donald Metzler, Xuanhui Wang, and Michael Bendersky. 2024. [Large language models are effective text rankers with pairwise ranking prompting](#). In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1504–1518, Mexico City, Mexico. Association for Computational Linguistics.
- Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. 2020. [Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters](#). In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 3505–3506.
- Devendra Sachan, Mike Lewis, Mandar Joshi, Armen Aghajanyan, Wen-tau Yih, Joelle Pineau, and Luke

- Zettlemoyer. 2022. [Improving passage retrieval with zero-shot question generation](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3781–3797.
- Rulin Shao, Rui Qiao, Varsha Kishore, Niklas Muennighoff, Xi Victoria Lin, Daniela Rus, Bryan Kian Hsiang Low, Sewon Min, Wen tau Yih, Pang Wei Koh, and Luke Zettlemoyer. 2025. [Reasonir: Training retrievers for reasoning tasks](#). *arXiv preprint arXiv:2504.20595*.
- Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. 2024. [Deepseekmath: Pushing the limits of mathematical reasoning in open language models](#). *arXiv preprint arXiv:2402.03300*.
- Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. 2025. [Hybridflow: A flexible and efficient rlhf framework](#). In *Proceedings of the Twentieth European Conference on Computer Systems*, pages 1279–1297.
- Taiwei Shi, Yiyang Wu, Linxin Song, Tianyi Zhou, and Jieyu Zhao. 2026. [Efficient reinforcement finetuning via adaptive curriculum learning](#). *Transactions on Machine Learning Research*.
- Seongwoong Shim, Myunsoo Kim, Jae Cho, and Byung-Jun Lee. 2026. [Beyond rag vs. long-context: Learning distraction-aware retrieval for efficient knowledge grounding](#). In *International Conference on Learning Representations*, volume 2026, pages 122532–122565.
- Mingyang Song, Mao Zheng, Zheng Li, Wenjie Yang, and Xuan Luo. 2025. [FastCuRL: Curriculum reinforcement learning with stage-wise context scaling for efficient training rl-like reasoning models](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 8856–8866, Suzhou, China. Association for Computational Linguistics.
- Hongjin Su, Howard Yen, Mengzhou Xia, Weijia Shi, Niklas Muennighoff, Han-yu Wang, Liu Haisu, Quan Shi, Zachary Siegel, Michael Tang, et al. 2025. [Bright: A realistic and challenging benchmark for reasoning-intensive retrieval](#). In *International Conference on Learning Representations*, volume 2025, pages 48941–48991.
- Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin, and Zhaochun Ren. 2023. [Is chatgpt good at search? investigating large language models as re-ranking agents](#). In *Proceedings of the 2023 conference on empirical methods in natural language processing*, pages 14918–14937.
- Hongye Tan, Xiaoyue Wang, Yu Ji, Ru Li, Xiaoli Li, Zhiwei Hu, Yunxiao Zhao, and Xiaoqi Han. 2021. [GCRC: A new challenging MRC dataset from Gaokao Chinese for explainable evaluation](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1319–1330, Online. Association for Computational Linguistics.
- Zecheng Tang, Ji Baibei, Juntao Li, Lijun Wu, Haijia Gui, and Min Zhang. 2026. [Revisiting long-context modeling from context denoising perspective](#). In *International Conference on Learning Representations*, volume 2026, pages 153734–153759.
- Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. 2021. [BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models](#). In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.
- Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. [Improving text embeddings with large language models](#). In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 11897–11916.
- Zhenteng Wang, Guofeng Cui, Kun Wan, and Wentian Zhao. 2025. [Dump: Automated distribution-level curriculum learning for rl-based llm post-training](#). *arXiv preprint arXiv:2504.09710*.
- William Webber, Alistair Moffat, and Justin Zobel. 2010. [A similarity measure for indefinite rankings](#). *ACM Transactions on Information Systems (TOIS)*, 28(4):1–38.
- Yiyang Wei, Tingyu Song, Siyue Zhang, and Yilun Zhao. 2026. [A survey of reasoning-intensive retrieval: Progress and challenges](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 42101–42122, San Diego, California, United States. Association for Computational Linguistics.
- Orion Weller, Kathryn Ricci, Eugene Yang, Andrew Yates, Dawn Lawrie, and Benjamin Van Durme. 2025. [Rank1: Test-time compute for reranking in information retrieval](#). *Preprint*, arXiv:2502.18418.
- Juan Wisznia, Cecilia Bolaños, Juan Tollo, Giovanni Franco Gabriel Marraffini, Agustín Andrés Gianolini, Noe Fabian Hsueh, and Luciano Del Corro. 2025. [Are optimal algorithms still optimal? rethinking sorting in LLM-based pairwise ranking with batching and caching](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1064–1072, Vienna, Austria. Association for Computational Linguistics.
- Hao Xu, Yunxiao Zhao, Jiayang Zhang, Zhiqiang Wang, and Ru Li. 2025. [LOG: A local-to-global optimization approach for retrieval-based explainable multi-hop question answering](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9085–9095, Abu Dhabi, UAE. Association for Computational Linguistics.

- Zhichao Yan, Jiapu Wang, Jiaoyan Chen, Xiaoli Li, Jiye Liang, Ru Li, and Jeff Z Pan. 2025. [Atomic fact decomposition helps attributed question answering](#). *IEEE Transactions on Knowledge and Data Engineering*.
- Zhichao Yan, Yunxiao Zhao, Jiapu Wang, Jiaoyan Chen, Xiaoli Li, Ru Li, and Jeff Z Pan. 2026. [Beyond factual accuracy: Evaluating global reasoning integrity in rag systems with logicscore](#). *arXiv preprint arXiv:2601.15050*.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. 2025a. [Qwen3 technical report](#). *arXiv preprint arXiv:2505.09388*.
- Eugene Yang, Andrew Yates, Kathryn Ricci, Orion Weller, Vivek Chari, Benjamin Van Durme, and Dawn Lawrie. 2025b. [Rank-k: Test-time reasoning for listwise reranking](#). *Preprint*, arXiv:2505.14432.
- Soyoung Yoon, Dongha Ahn, Youngwon Lee, Minkyu Jung, HyungJoo Jang, and Seung-won Hwang. 2025. [RoToR: Towards more reliable responses for order-invariant inputs](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 18739–18760, Vienna, Austria. Association for Computational Linguistics.
- Soyoung Yoon, Eunbi Choi, Jiyeon Kim, Hyeongu Yun, Yireun Kim, and Seung-won Hwang. 2024. [ListT5: Listwise reranking with fusion-in-decoder improves zero-shot retrieval](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2287–2308, Bangkok, Thailand. Association for Computational Linguistics.
- Soyoung Yoon, Gyuwan Kim, Gyu-Hwung Cho, et al. 2026. [Acurank: Uncertainty-aware adaptive computation for listwise reranking](#). *Advances in Neural Information Processing Systems*, 38:26388–26417.
- Le Zhang, Bo Wang, Xipeng Qiu, Siva Reddy, and Aishwarya Agrawal. 2025a. [REARANK: Reasoning re-ranking agent via reinforcement learning](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 2458–2471, Suzhou, China. Association for Computational Linguistics.
- Xiangxu Zhang, Lei Li, Xiao Zhou, and Zheng Liu. 2025b. [R2med: A benchmark for reasoning-driven medical retrieval](#). *arXiv preprint arXiv:2505.14558*.
- Yunxiao Zhao, Zhiqiang Wang, Xiaoli Li, Jiye Liang, and Ru Li. 2024. [AGR: Reinforced causal agent-guided self-explaining rationalization](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 510–518, Bangkok, Thailand. Association for Computational Linguistics.
- Yunxiao Zhao, Zhiqiang Wang, Xingtong Yu, Xiaoli Li, Jiye Liang, and Ru Li. 2026. [Learnable game-theoretic policy optimization for data-centric self-explanation rationalization](#). *IEEE Transactions on Knowledge and Data Engineering*, 38(2):1159–1173.
- Yunxiao Zhao, Hao Xu, Zhiqiang Wang, Xiaoli Li, Jiye Liang, and Ru Li. 2025. [Explaining black-box language models with knowledge probing systems: A post-hoc explanation perspective](#). In *International Conference on Database Systems for Advanced Applications*, pages 220–235. Springer.
- Lianmin Zheng, Liangsheng Yin, Zhiqiang Xie, Chuyue Livia Sun, Jeff Huang, Cody Hao Yu, Shiyi Cao, Christos Kozyrakis, Ion Stoica, Joseph E Gonzalez, et al. 2024a. [Sglang: Efficient execution of structured language model programs](#). *Advances in neural information processing systems*, 37:62557–62583.
- Yaowei Zheng, Richong Zhang, Junhao Zhang, Yanhan Ye, and Zheyang Luo. 2024b. [LlamaFactory: Unified efficient fine-tuning of 100+ language models](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 400–410, Bangkok, Thailand. Association for Computational Linguistics.
- Yutao Zhu, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Haonan Chen, Zheng Liu, Zhicheng Dou, and Ji-Rong Wen. 2025. [Large language models for information retrieval: A survey](#). *ACM Transactions on Information Systems*, 44(1):1–54.
- Shengyao Zhuang, Xueguang Ma, Zheng Yao, Shuai Wang, Bevan Koopman, Jimmy Lin, and Guido Zuccon. 2026. [Rank-r1: Enhancing reasoning in llm-based document rerankers via reinforcement learning](#). In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 4419–4425.

Appendix

A Benchmarks

In our experiments, we utilize three IR benchmarks for evaluation: BRIGHT (Su et al., 2025), R2MED (Zhang et al., 2025b) and BEIR (Thakur et al., 2021). These benchmarks are used to evaluate the performance of retriever and reranker across different scenarios, enabling a more comprehensive assessment. A detailed introduction to each benchmark is provided below:

BRIGHT The BRIGHT benchmark is designed to challenge information retrieval systems with complex queries that require deep reasoning capabilities. Unlike traditional benchmarks that primarily focus on simple keyword matching or semantic similarity, BRIGHT consists of 1,384 queries spanning diverse domains, including economics, psychology, mathematics, and coding. These queries are derived from both naturally occurring data and carefully curated human data, making the benchmark a more comprehensive and rigorous evaluation for advanced retrieval models. Since identifying relevant documents requires substantial reasoning, BRIGHT imposes higher demands on retrieval systems. It requires models not only to perform surface-level matching, but also to demonstrate deeper cognitive and reasoning abilities.

R2MED R2MED is a specialized benchmark for medical information retrieval, emphasizing the importance of reasoning in retrieving relevant medical information. It features 876 queries that cover three main tasks from 8 datasets: Q&A reference retrieval, clinical evidence retrieval, and clinical case retrieval. These tasks span a wide array of medical scenarios and body systems, highlighting the necessity for retrieval systems to understand and process medical information accurately. R2MED underscores the critical role reasoning plays in medical decision-making, challenging retrieval systems to move beyond simple lexical or semantic matching to support complex clinical judgments and evidence-based practice.

BEIR BEIR is a comprehensive benchmark designed to evaluate the performance of information retrieval systems across 18 diverse datasets. These datasets encompass a variety of tasks, including fact-checking, question answering, and more, allowing for a thorough assessment of a model’s generalization capabilities. BEIR is particularly

focused on the ability of retrieval systems to adapt across different domains and tasks, emphasizing the need for a balance between performance and computational efficiency. By encouraging the development of robust and adaptable retrieval systems, BEIR serves as a pivotal benchmark in the advancement of information retrieval technologies.

B Baselines

The baselines we used for comparison are as follows, each selected for its unique strengths and capabilities in specific retrieval or reranking tasks:

B.1 Retrievers

ReasonIR (Shao et al., 2025): ReasonIR is a reasoning-intensive retriever training on reasoning-intensive synthetic data based on LLAMA3.1-8B (Grattafiori et al., 2024). This baseline is particularly adept at handling complex queries that require in-depth reasoning, making it an ideal choice as our initial retriever on the BRIGHT benchmark. In the main experiments, we use the GPT-4-rewritten queries provided in the test set for retrieval, as they achieve better retrieval performance, while the original test queries are used for reranking.

E5-Mistral-7B-Instruct (Wang et al., 2024): E5-Mistral-7B-Instruct² is fine-tuned on a mixture of multilingual datasets and has some multilingual capability. In the main experiments, we use it as our initial retriever on the R2MED benchmark and use the original test query for retrieval and reranking.

ReasonEmbed (Chen et al., 2026): ReasonEmbed is a text embedding model specifically designed for reasoning-intensive document retrieval. It improves retrieval performance by leveraging reasoning-aware synthetic data generation and an adaptive training strategy that emphasizes harder reasoning samples. We employ ReasonEmbed (4B) for the initial retrieval stage on the R2MED benchmark to investigate the performance of models under different retrievers.

B.2 Non-Reasoning Rerankers

RankLLaMA (Ma et al., 2024): A pointwise reranker built upon the LLaMA (Grattafiori et al., 2024) model and fine-tuned through Supervised Fine-Tuning (SFT) on the MS MARCO (Bajaj et al., 2018) datasets.

²<https://huggingface.co/intfloat/e5-mistral-7b-instruct>

RankZephyr (Pradeep et al., 2023): A listwise reranker is distilled from the powerful models GPT-3.5 and GPT-4 (Achiam et al., 2023), utilizing a sophisticated dual-phase training framework. During inference, it employs a sliding window strategy for listwise reranking.

FIRST (Gangi Reddy et al., 2024): An efficient listwise reranking method that improves decoding speed by optimizing logit generation within a single window, though it still relies on sequential sliding window strategy to handle long candidate lists.

B.3 Reasoning Rerankers

ERank (Cai et al., 2026): A pointwise reranker that fuses SFT with Reinforcement Learning (RL) to enhance both the effectiveness and efficiency of text reranking.

JudgeRank (Niu et al., 2024): A pointwise reranker designed for reasoning-intensive retrieval tasks. By simulating the cognitive process humans use to assess document relevance, it performs step-by-step analysis of queries and documents to improve reranking performance.

Rank1 (Weller et al., 2025): A pointwise reranker that leverages test-time compute and reasoning trace distillation from reasoning-oriented large language models (LLMs) to enhance reranking performance, generalization ability, and interpretability.

Retro* (Lan et al., 2026): A reasoning-enhanced pointwise reranker that leverages rubric-based relevance scoring, test-time scaling, and reinforcement learning to improve fine-grained relevance estimation, interpretability, and retrieval performance on complex retrieval tasks.

Rank-R1 (Zhuang et al., 2026): A reasoning-enhanced listwise reranker that employs reinforcement learning to improve the reasoning and relevance assessment capabilities of LLM-based rerankers, thereby enhancing ranking effectiveness and explainability for complex queries.

REARANK (Zhang et al., 2025a): A reasoning-enhanced listwise reranker that explicitly performs reasoning before reranking, leveraging reinforcement learning and data augmentation to improve reranking performance and interpretability on complex retrieval tasks.

GroupRank (Long et al., 2026): A groupwise reranker that balances the efficiency of pointwise methods and the contextual awareness of listwise methods, leveraging supervised fine-tuning and reinforcement learning to improve reranking performance and efficiency.

TourRank (Chen et al., 2025b): A tournament-inspired tournament-style reranker that leverages a multi-stage grouping strategy and a points-based ensemble system to overcome input length limitations and mitigate position bias.

ReasonRank (Liu et al., 2026a): A reasoning-enhanced listwise reranker that employs a two-stage training strategy (SFT on synthetic reasoning data followed by RL) to empower the model with strong logical reasoning abilities.

C Implementation Details

To provide a clearer and more comprehensive understanding of our CuRLRank, we present detailed implementation details for each key component, including Difficulty-Aware Dataset Construction, Supervised Fine-Tuning (SFT), Adaptive Curriculum Reinforcement Learning (CuRL), and the Iterative Filtering Ranking Strategy (IFRS). We also provide additional experiments to further demonstrate the effectiveness and superiority of our CuRLRank.

C.1 Details about Difficulty-Aware Dataset Construction

We construct our training data from the synthetic dataset provided by ReasonEmbed³. Specifically, we use ReasonEmbed-Qwen3-8B (Chen et al., 2026) as the retriever and employ DeepSeek-R1 (Guo et al., 2025) as the teacher model for data annotation. Each sample consists of a query and a list of documents, where the document list contains 3 positive documents and 7 negative documents. For each dataset, we define a relevance rubric Γ and design a prompt template that instructs the model to generate a reasoning process r , a list of relevance scores S , and an initial difficulty level h . Each relevance score s is an integer between 0 and 100, and the initial difficulty level h is an integer between 1 and 5. The prompt template will be provided in Table 13. The relevance definitions for the three benchmarks are provided in Table 14, 15, and 16.

During the difficulty adjustment stage, we use Qwen3-4B-Instruct-2507⁴ to compute the token length of each sample’s reasoning process. We then adjust the initial difficulty level h^{model} by incorporating both the length-based difficulty bias δ^{length} and score-based difficulty bias δ^{score} , resulting in

³<https://huggingface.co/datasets/hanhainebola/reason-embed-data>

⁴<https://huggingface.co/Qwen/Qwen3-4B-Instruct-2507>

a final difficulty level h for each sample. The final difficulty h is a floating-point value ranging from 0 to 5.

To ensure the accuracy of the difficulty levels, we further conducted human annotation for verification. We randomly sampled 40 examples from each dataset in BRIGHT, resulting in 480 examples in total. Four graduate students who are specialized in NLP independently annotated these examples, and we computed the average difficulty. Table 5 shows strong alignment between human-annotated difficulty and the final difficulty levels. Across all four metrics, the top-3 values are consistently concentrated on the Leet., AoPS, and TheoT. datasets, further validating the reliability of our difficulty levels.

Finally, we construct 14,145 training samples in total, where 6,000 samples are used for SFT and 8,145 samples are used for CuRL.

C.2 Details about Supervised Fine-Tuning

We use LLaMA-Factory (Zheng et al., 2024b) to train two different base models: Qwen3-4B-Instruct-2507 and Qwen2.5-7B-Instruct⁵. We perform full fine-tuning on both models, setting the learning rate to 1×10^{-5} , with per GPU batch size of 1 and gradient accumulation steps of 8. We adopt DeepSpeed ZeRO-3 (Rasley et al., 2020) and FlashAttention-2 (Dao, 2024) to accelerate training. We use BF16 mixed precision and set the maximum input sequence length to 8192. Both models are trained for 5 epochs. For hardware, we train Qwen3-4B-Instruct-2507 on $4 \times$ NVIDIA RTX 6000 Ada 48GB GPUs, and Qwen2.5-7B-Instruct on $4 \times$ NVIDIA A800 80GB GPUs. Detailed hyperparameter settings are provided in Table 6.

C.3 Details about Adaptive Curriculum Reinforcement Learning

We perform adaptive curriculum reinforcement learning on the model obtained via supervised fine-tuning. Built upon the verl⁶ framework, we train our CuRLRank (4B and 7B) using the GRPO algorithm (Shao et al., 2024). During training, we employ the vLLM⁷ inference engine to accelerate generation and inference.

⁵<https://huggingface.co/Qwen/Qwen2.5-7B-Instruct>

⁶<https://github.com/verl-project/verl>

⁷<https://github.com/vllm-project/vllm>

We use the GRPO reinforcement learning algorithm (Shao et al., 2024) to optimize the model. Both the policy model π_θ and the reference model π_{ref} are initialized from the supervised fine-tuning reranker. For each input x , we sample a group of outputs $\{y_1, \dots, y_{|G|}\}$ from the old policy $\pi_{\theta_{\text{old}}}$. Each sequence y_i receives a reward R_i computed using the score-based reward function, which is further normalized within the group to obtain the relative advantage $\hat{A}_i$. The optimization objective maximizes a clipped surrogate loss regularized by a KL divergence penalty $\mathbb{D}_{\text{KL}}$:

$$\begin{aligned} \mathcal{J}_{\text{GRPO}}(\theta) = & \frac{1}{|G|} \sum_{i=1}^{|G|} \frac{1}{|y_i|} \sum_{t=1}^{|y_i|} \min \left(\rho_{i,t}(\theta) \hat{A}_i, \right. \\ & \left. \text{clip}(\rho_{i,t}(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_i \right) \\ & - \beta \mathbb{D}_{\text{KL}}(\pi_\theta \parallel \pi_{\text{ref}}), \end{aligned} \quad (15)$$

where ϵ and β are hyperparameters and $\rho_{i,t}(\theta)$ is the token probability ratio defined as:

$$\rho_{i,t}(\theta) = \frac{\pi_\theta(y_{i,t} \mid x, y_{i,<t})}{\pi_{\text{old}}(y_{i,t} \mid x, y_{i,<t})}. \quad (16)$$

The KL divergence term $\mathbb{D}_{\text{KL}}(\pi_\theta \parallel \pi_{\text{ref}})$ constrains the policy model toward the reference model to stabilize training and prevent excessive policy drift. Through this iterative optimization process, CuRLRank progressively improves its ranking capability via adaptive curriculum reinforcement learning while preserving the foundational ranking ability acquired during supervised fine-tuning.

C.4 Details about Iterative Filtering Ranking Strategy

We employ IFRS to rerank the top-100 candidate documents. Specifically, both the score smoothing weight β and the retention ratio γ are set to 0.5, while the window size is set to 10. The maximum length for both query and document is set to 768 tokens. In each iteration, half of the candidate documents are filtered out. The iterative process terminates when the number of remaining documents is less than or equal to 10, and the top-10 documents with the highest scores are regarded as the most relevant documents.

We note that our IFRS shares certain similarities with the previously proposed TourRank (Chen et al., 2025b). Compared with TourRank, our IFRS offers three main advantages: (1) We adopt fine-grained relevance scores ranging from 0 to 100,

Metric	StackExchange							Coding		Theorems		
	Econ.	Earth.	Rob.	Bio.	Psy.	Stack.	Sus.	Leet.	Pony.	AoPS	TheoQ.	TheoT.
Human Level	3.13	2.63	2.26	2.45	2.34	3.22	2.81	3.75	2.34	3.96	3.46	3.53
Final Level	2.98	2.91	2.39	2.58	2.17	3.16	2.94	3.67	2.51	3.84	3.33	3.64
Within-0.5 (%)	82.42	77.65	74.31	76.08	75.47	83.19	79.73	82.56	81.95	85.21	81.67	85.38
Inconsistency (%)	10.76	8.39	7.87	8.25	7.98	11.63	9.19	16.55	8.16	19.28	14.84	17.77

Table 5: Comparison of human annotation, final difficulty level, and inconsistency. Within-0.5 (%) denotes the percentage of samples whose deviation between human difficulty and final difficulty is within 0.5. Inconsistency (%) measures the discrepancy between the model-assigned scores and the original relevance annotations. The top-3 values are highlighted in **bold**.

Hyperparameter	Value
SFT	
cutoff_len	8192
per_device_train_batch_size	1
gradient_accumulation_steps	8
learning_rate	1×10^{-5}
num_train_epochs	5
CuRL	
train_batch_size	32
max_prompt_length	8192
max_response_length	1536
learning_rate	1×10^{-5}
kl_loss_coef	0.001
rollout_n	8
ppo_micro_batch_size	1
ppo_mini_batch_size	8
num_train_epochs	4
scaling factor α	0.05
temperature coefficient τ	20
difficulty step size η	0.02
IFRS	
smoothing weight β	0.5
retention ratio γ	0.5

Table 6: Hyperparameter settings for SFT, CuRL, and IFRS stages.

Method	Parallel	Cross-D.	Complexity
Pointwise	✓		$\mathcal{O}(N)$
Pairwise (all-pair)	✓	✓	$\mathcal{O}(N^2)$
Pairwise (heapsort)		✓	$\mathcal{O}(k \cdot \log_2 N)$
Pairwise (bubblesort)		✓	$\mathcal{O}(k \cdot N)$
Listwise (SW)		✓	$\mathcal{O}(N/s)$
Listwise (IFRS)	✓	✓	$\mathcal{O}(2N/w)$

Table 7: Comparison of different reranking strategies in terms of parallel and cross-document (Cross-D.) ranking capability. Here, SW denotes the sliding-window strategy. N is the number of documents, k is the number of most relevant documents, w is the window size, and s the sliding step.

Model	Avg.	Increase
ReasonIR (8B)	30.59	-
Retro* (7B)	35.61	-
Retro* (7B) + IFRS	36.73	+1.12
GroupRank (7B)	33.49	-
GroupRank (7B) + IFRS	36.34	+2.85
CuRLRank (7B) - IFRS	35.79	-
CuRLRank (7B)	37.85	+2.06

Table 8: NDCG@10 of different rerankers with and without IFRS on the BRIGHT benchmark. All models rerank the top-100 documents retrieved by ReasonIR (8B).

which provide better explainability than the binary relevance scores (0/1) used in TourRank; (2) By accumulating historical scores across iterations (Zhao et al., 2024), IFRS effectively integrates previous ranking information to improve robustness throughout the multi-round ranking process; (3) IFRS incorporates a global fairness mechanism combining local scoring with global filtering, ensuring that high-scoring documents are not prematurely eliminated because of unfavorable initial grouping.

Compared with the widely used sliding window ranking strategy, our method additionally supports parallel computation and therefore achieves higher computational efficiency. Table 7 presents a comparison between IFRS and other reranking strategies. It can be observed that our IFRS supports both parallel processing and cross-document ranking while maintaining low time complexity. Table 8 reports the NDCG@10 (%) results of IFRS with different rerankers, which demonstrates that IFRS can significantly improve ranking performance and has broad generality.

D More Experiments for CuRLRank

D.1 Different Backbones and Retrievers

To validate the generalizability and effectiveness of our proposed CuRLRank framework, we fur-

Model	StackExchange							Coding		Theorems			Avg.
	Econ.	Earth.	Rob.	Bio.	Psy.	Stack.	Sus.	Leet.	Pony.	AoPS	TheoQ.	TheoT.	
ReasonIR (8B)	32.65	43.00	20.82	43.49	39.57	30.96	27.34	31.69	19.55	7.37	33.93	36.68	30.59
Rank1 (32B)	25.43	37.97	17.14	42.43	34.85	23.82	31.18	12.20	40.98	4.79	29.28	39.95	28.34
Retro* (32B)	41.39	56.27	35.12	<u>58.33</u>	49.64	37.81	44.19	14.63	26.94	<u>10.30</u>	36.82	45.74	<u>38.10</u>
GroupRank (32B)	39.18	50.01	<u>35.13</u>	57.53	51.42	39.00	42.63	23.22	21.45	10.12	37.66	47.09	37.87
ReasonRank (32B)	36.64	48.90	33.88	58.17	<u>53.27</u>	38.68	45.97	25.78	20.93	9.41	<u>38.46</u>	46.28	38.03
CuRLRank (32B)	<u>40.12</u>	<u>52.76</u>	37.74	59.46	54.88	40.35	44.08	<u>25.49</u>	24.61	11.39	40.24	49.63	40.06

Table 9: NDCG@10 results (%) of different 32B rerankers on the BRIGHT benchmark. All models rerank the top-100 documents retrieved by ReasonIR (8B). The top two results are highlighted in **bold** and underlined, respectively.

Model	Covid	DBPedia	SciFact	NFCorpus	Signal	Robust04	News	Avg.
BM25	59.47	31.80	67.89	33.75	33.05	40.70	39.52	43.74
Non-Reasoning Reranker								
RankLLaMA (7B)	84.67	<u>48.29</u>	73.21	30.28	33.73	56.64	51.01	53.98
RankZephyr (7B)	82.92	44.42	75.42	38.26	31.41	53.73	52.80	54.14
FIRST (7B)	78.63	49.78	74.64	37.44	32.59	54.63	51.86	54.22
Reasoning Reranker								
ERank (4B)	79.68	39.17	73.14	35.87	27.93	56.39	46.92	51.30
JudgeRank (7B)	78.23	41.56	72.31	35.43	23.16	55.45	44.33	50.07
Rank1 (7B)	79.04	35.79	73.32	37.52	25.41	57.11	47.67	50.84
Retro* (7B)	83.56	41.87	76.31	39.57	29.33	64.70	49.53	54.98
Rank-R1 (7B)	83.71	42.27	72.16	38.94	<u>33.08</u>	54.46	50.60	53.60
REARANK (7B)	81.28	44.99	75.02	38.83	32.69	56.16	51.04	54.29
GroupRank (7B)	79.82	45.21	71.04	37.18	30.01	56.94	51.23	53.06
TourRank (7B) [†]	82.56	44.64	72.17	37.99	30.83	57.87	51.46	53.93
ReasonRank (7B)	82.89	46.21	75.81	<u>39.42</u>	31.61	55.74	50.87	54.65
CuRLRank (4B)	<u>84.24</u>	43.75	<u>80.12</u>	37.17	29.92	<u>65.31</u>	<u>52.20</u>	<u>56.10</u>
CuRLRank (7B)	83.17	45.84	80.99	38.26	28.75	66.20	51.14	56.34

Table 10: NDCG@10 on 7 datasets in the BEIR. All models rerank the top-100 documents retrieved by BM25. The top two results are highlighted in **bold** and underlined, respectively. [†] denotes results reproduced by us, as the models are not publicly available. Pointwise and Listwise rerankers are indicated by gray and light blue background respectively, while our **CuRLRank** is highlighted in blue background.

ther conduct experiments with different backbone models and retrievers.

For backbone models, in addition to Qwen3-4B-Instruct-2507 and Qwen2.5-7B-Instruct used in our main experiments, we further adopt Qwen2.5-32B-Instruct⁸ to evaluate performance at a larger model scale, and Llama-3.1-8B-Instruct⁹ to assess generalizability across different model families. For Qwen3-4B-Instruct-2507 and Qwen2.5-7B-Instruct, we perform full-parameter fine-tuning. For Qwen2.5-32B-Instruct, we adopt LoRA for efficient fine-tuning, with the learning rate set to 1×10^{-5} and both the LoRA rank and alpha set to 32. Detailed training hyperparameters are provided in Table 6.

For retrievers, we employ different retrievers

⁸<https://huggingface.co/Qwen/Qwen2.5-32B-Instruct>

⁹<https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct>

from those used in the main experiments. On the BRIGHT benchmark, we use BGE-Reasoner-Rewritten-Query-0821¹⁰ for multiple query rewrite and adopt the traditional sparse retriever BM25 as the initial retriever¹¹. On the R2MED benchmark, we directly use the original queries and employ reason-embed-qwen3-4b-0928¹² (abbreviated as ReasonEmbed (4B)) as the initial retriever. For each query, we retrieve the top-100 documents for reranking. In addition, we also report a variant trained only with supervised fine-tuning (SFT), denoted as **CuRLRank_{Only-SFT}**.

As shown in Table 11 and 12, CuRLRank consistently and significantly outperforms all baseline

¹⁰<https://huggingface.co/datasets/cfli/reasoner-rewritten-query-0821>

¹¹The initial retrieval results can be found at: <https://huggingface.co/datasets/hanhaineibula/bright-search-results>

¹²<https://huggingface.co/hanhaineibula/reason-embed-qwen3-4b-0928>

across different backbone models and retrievers, validating the effectiveness of our CuRL and IFRS. Moreover, compared with CuRLRank_{Only-SFT}, CuRLRank achieves substantial improvements on the average metrics, further demonstrating the effectiveness of the score-based reward function and adaptive curriculum sampling.

Notably, CuRLRank (32B) achieves the best average NDCG@10 scores of 42.05 and 49.39 on BRIGHT and R2MED, respectively. Table 9 further compares CuRLRank with mainstream rerankers, where CuRLRank (32B) consistently achieves the best overall performance. These results demonstrate that CuRLRank remains effective when scaled to larger backbone models, highlighting its strong generalizability across different model scales.

Interestingly, on several BRIGHT subsets, although the sparse retriever BM25 produces weaker initial retrieval results than the dense retriever ReasonIR (8B), such as on TheoQ. (29.49 vs 33.93) and TheoT. (25.13 vs 36.68), the reranked results produced by CuRLRank substantially surpass those based on ReasonIR (8B). This observation suggests two important insights: (1) the choice of the initial retriever is crucial for the final ranking performance, as different retrievers focus on different retrieval characteristics, and hybrid retrieval may therefore lead to better overall performance; (2) CuRLRank is capable of compensating for the limitations of BM25 by more accurately identifying the most relevant documents during ranking.

In addition, we further compare CuRLRank with the teacher model on both benchmarks. The results show that DeepSeek-R1 (671B) (Guo et al., 2025) only slightly outperforms our best CuRLRank (32B) by approximately 1 point. This demonstrates that our CuRLRank can effectively learn the reasoning and ranking capabilities of the teacher model. However, the labels generated by the teacher model are not always fully reliable. Therefore, during reinforcement learning, the model should not overly rely on the annotations provided by the teacher model, and need to further optimization through reward function and curriculum learning.

D.2 Details about Contextual Sensitivity

Analysis

We evaluate the performance and robustness of different rerankers on the BRIGHT and R2MED benchmarks. Due to computational constraints and

the limited number of positive samples in some datasets, we construct a dedicated evaluation set for robustness analysis. We select several mainstream listwise rerankers as baselines to compare with CuRLRank, including REARANK, GroupRank, TourRank and ReasonRank. Detailed introductions to these rerankers are provided in Appendix B. To ensure a fair comparison, all rerankers are evaluated using 7B-scale models. For TourRank, since the model has not been publicly released, we use Qwen2.5-7B-Instruct as the backbone model and train it on a subset of the ReasonEmbed (Chen et al., 2026) training data.

To ensure data diversity, we randomly sample M instances from each dataset in the two benchmarks. Each instance consists of one query and $(m + 2n)$ candidate documents, including m positive documents and $2n$ negative documents. In total, we construct an evaluation dataset containing N instances.

First, we use the m positive documents together with half of the negative documents (n negative documents) as the candidate set, and let the model generate the initial ranking result L_I . (1) Order Perturbation: We randomly shuffle all candidate documents and let the model generate the order-perturbed ranking L_O . The overlap between L_I and L_O reflects the sensitivity to document order perturbation. (2) Content Perturbation: We combine the positive documents with the other half of the negative documents to form a new candidate set, and let the model generate the content-perturbed ranking L_C . The overlap between L_I and L_C reflects the robustness of the reranker under content perturbation induced by unseen distractor documents.

Then, we use Rank-Biased Overlap (RBO) (Webber et al., 2010) to measure the similarity between the top- k documents in the initial ranking L_I and the perturbed ranking L_P ($L_P \in \{L_O, L_C\}$). Then, we define the robustness score Rob to measure the ranking stability which is the mean of $RBO_k(L_I, L_P)$ across all N test instances:

$$RBO_k(L_I, L_P) = (1-p) \sum_{d=1}^k p^{d-1} \frac{|L_{I;d} \cap L_{P;d}|}{d}, \quad (17)$$

$$Rob = \frac{1}{N} \sum_{i=1}^N RBO_k(L_I^{(i)}, L_P^{(i)}). \quad (18)$$

The $L_{I;d}$ and $L_{P;d}$ are the top- d documents of each list, and p controls weight decay. In our ex-

periments, we specify these hyperparameters as $M = 20$, $N = 400$, $m = 10$, $n = 90$, $k = 10$ and $p = 0.9$.

D.3 Details about Hyperparameter Sensitivity Analysis

Since the relevance scores lie in a relatively large range of $[0, 100]$, we introduce scaling factors to align both sigmoid and softmax operations with their effective operating regimes, where the non-saturated (approximately linear) region of the sigmoid function is typically within $[-5, 5]$. Specifically, we set scaling factor $\alpha = 0.05$ to rescale pairwise score differences into the non-saturated region of the sigmoid function, ensuring informative gradients for hard pairs. Similarly, we set temperature coefficient $\tau = 20$ to normalize negative scores before softmax, preventing probability collapse while preserving relative hardness among negative samples.

We randomly select 240 samples from the training dataset to compute initial average reward $\bar{R}_0$ using our predefined reward function (Eq 10). The resulting initial average reward is 0.685. To enable a broader difficulty range for the model, we set the initial target difficulty $\bar{H}_0$ to 0. The minimum and maximum difficulty levels are $H_{\min} = 0$ and $H_{\max} = 5$, respectively, and the difficulty H_t is defined as a continuous value in the range $[0, 5]$.

The difficulty step size η is determined based on the total number of training samples, the training batch size, and the difficulty range. Our goal is to ensure that the model gradually traverses all difficulty levels in the training data. Specifically, we assume the model increases by one difficulty step at each training step, and compute η as follows:

$$\eta = \frac{(H_{\max} - H_{\min})B}{N}, \quad (19)$$

where N denotes the total number of training samples, $H_{\min}$ and $H_{\max}$ are the minimum and maximum difficulty levels, respectively, and B is the training batch size. The specific values are provided in Table 6.

D.4 Detailed Results on BEIR Benchmark

Following previous settings (Sun et al., 2023; Liu et al., 2026a), we present comparison results on seven datasets from the traditional IR benchmark BEIR (Thakur et al., 2021). We use BM25 as the retriever and rerank the top-100 retrieved documents. Meanwhile, we evaluate CuRLRank (4B and 7B)

against the same baselines as in the main experiments, with results reported in Table 10. CuRLRank (4B and 7B) achieves the best average performance among all methods. In particular, CuRLRank (4B) outperforms the previous best baseline Retro* (Lan et al., 2026) by 1.12, while CuRLRank (7B) further improves over it by 1.36, demonstrating the effectiveness of CuRLRank.

E Case Studies of CuRLRank

To better understand the reasoning process of our CuRLRank, we present the cases of CuRLRank (4B) on several BRIGHT benchmark datasets, including Biology (Table 17), Earth Science (Table 18), Economics (Table 19), Leetcode (Table 20), Pony (Table 21) and TheoremQA-Questions (Table 22). Notably, the *Difficulty Annotation* step is only introduced during the data annotation stage. During both model training and evaluation, this step is removed and only the first three steps are retained. The purpose is to encourage the student model to focus on generating accurate relevance score lists rather than performing difficulty annotation.

F Use of AI Assistants

We use ChatGPT to improve the presentations of this paper¹³.

¹³<https://chatgpt.com/>

Model	StackExchange							Coding		Theorems			Avg.
	Econ.	Earth.	Rob.	Bio.	Psy.	Stack.	Sus.	Leet.	Pony.	AoPS	TheoQ.	TheoT.	
BM25	23.41	55.18	24.82	58.34	38.31	26.65	27.95	23.83	26.81	5.62	29.49	25.13	30.46
ERank (4B)	29.79	49.04	28.43	48.72	43.12	31.52	38.17	28.54	24.77	8.46	36.75	47.32	34.55
Rank1 (7B)	23.94	38.43	28.93	51.26	31.73	24.10	38.42	19.33	31.26	6.41	38.46	45.23	31.46
Retro* (7B)	31.02	57.18	35.78	55.54	48.47	34.80	40.60	26.89	33.36	7.68	36.32	50.78	38.20
Rank-R1 (7B)	23.13	32.16	26.77	41.24	33.09	15.31	33.48	23.10	24.96	5.91	35.74	48.28	28.60
REARANK (7B)	26.78	38.01	29.34	43.66	39.16	25.65	33.62	<u>28.92</u>	31.24	6.43	39.67	46.21	32.39
ReasonRank (7B)	31.22	52.72	34.31	57.74	46.65	30.69	39.03	24.44	37.05	10.00	40.24	57.71	38.48
<i>Using Llama-3.1-8B-Instruct as the backbone</i>													
CuRLRank _{Only-SFT} (8B)	30.18	43.66	26.58	53.10	45.03	26.12	38.77	22.15	34.41	7.09	35.19	46.66	34.08
CuRLRank (8B)	35.59	46.79	33.37	56.43	<u>52.89</u>	28.81	43.68	22.93	39.92	8.75	39.58	51.02	38.31
<i>Using Qwen3-4B-Instruct-2507 as the backbone</i>													
CuRLRank _{Only-SFT} (4B)	34.76	45.70	35.84	51.81	43.27	32.73	41.99	25.75	35.28	7.67	38.12	45.38	36.53
CuRLRank (4B)	37.10	50.06	37.82	57.87	47.48	35.32	44.83	23.06	40.72	12.79	40.32	52.14	39.96
<i>Using Qwen2.5-7B-Instruct as the backbone</i>													
CuRLRank _{Only-SFT} (7B)	31.24	48.55	33.62	52.40	44.77	29.96	39.92	29.78	32.54	8.96	40.65	47.43	36.65
CuRLRank (7B)	36.93	53.46	36.71	58.96	48.12	34.08	43.83	25.14	38.71	9.43	<u>41.20</u>	53.91	40.04
<i>Using Qwen2.5-32B-Instruct as the backbone</i>													
CuRLRank _{Only-SFT} (32B)	34.98	48.65	35.23	55.19	46.68	33.99	40.53	28.46	37.77	7.20	39.36	53.62	38.47
CuRLRank (32B)	39.46	52.33	<u>37.64</u>	<u>60.71</u>	51.93	<u>35.01</u>	<u>45.46</u>	25.18	43.72	10.38	43.27	59.49	<u>42.05</u>
<i>Teacher Model</i>													
DeepSeek-R1 (671B)	<u>38.88</u>	<u>54.52</u>	37.18	63.57	54.16	34.82	47.23	26.44	<u>42.95</u>	<u>11.16</u>	40.80	61.02	42.73

Table 11: NDCG@10 on the BRIGHT benchmark. All models rerank the top-100 documents retrieved by BM25 (using the multiple rewritten queries from BGE-Reasoner-Rewritten-Query-0821). The top two results are highlighted in **bold** and underlined. Pointwise and Listwise rerankers are indicated by gray and light blue background respectively, while our CuRLRank is highlighted in blue background.

Model	Q&A Reference			Clinical Evidence			Clinical Case		Avg.
	Biology	Bioin.	MedS.	MedE.	MedD.	PMCT.	PMCC.	IYiC.	
ReasonEmbed (4B)	49.72	63.77	60.30	21.41	25.12	44.61	30.11	22.47	39.69
ERank (4B)	41.83	63.18	63.74	23.44	32.15	30.23	36.94	19.42	38.87
Rank1 (7B)	54.26	64.63	61.13	22.66	36.17	51.48	31.25	18.37	42.49
Retro* (7B)	53.52	61.87	62.66	21.95	36.64	58.76	37.32	21.75	44.31
Rank-R1 (7B)	55.21	61.04	56.38	23.17	38.13	54.81	29.66	25.20	42.95
REARANK (7B)	54.46	62.83	57.99	22.85	36.59	54.72	34.98	24.16	43.57
ReasonRank (7B)	55.67	62.04	61.83	25.03	38.55	55.85	36.26	28.19	45.43
<i>Using Llama-3.1-8B-Instruct as the backbone</i>									
CuRLRank _{Only-SFT} (8B)	48.26	55.14	57.35	17.26	36.27	54.03	36.20	24.87	41.17
CuRLRank (8B)	50.18	66.39	65.81	25.19	38.44	60.78	39.57	27.15	46.69
<i>Using Qwen3-4B-Instruct-2507 as the backbone</i>									
CuRLRank _{Only-SFT} (4B)	48.37	56.10	56.42	24.10	36.00	56.26	38.55	30.97	43.35
CuRLRank (4B)	52.61	64.35	61.92	28.75	39.37	63.87	<u>43.93</u>	28.75	47.94
<i>Using Qwen2.5-7B-Instruct as the backbone</i>									
CuRLRank _{Only-SFT} (7B)	49.15	57.32	55.56	28.30	37.91	57.22	40.24	27.03	44.09
CuRLRank (7B)	54.23	63.97	65.43	26.66	<u>41.28</u>	64.49	41.81	29.92	48.47
<i>Using Qwen2.5-32B-Instruct as the backbone</i>									
CuRLRank _{Only-SFT} (32B)	52.25	60.94	59.31	26.47	40.56	62.53	42.26	<u>29.36</u>	46.71
CuRLRank (32B)	<u>55.97</u>	<u>66.81</u>	<u>66.74</u>	28.69	38.22	<u>66.70</u>	44.85	27.14	<u>49.39</u>
<i>Teacher Model</i>									
DeepSeek-R1 (671B)	56.14	68.22	67.39	26.48	43.45	69.73	43.81	28.01	50.40

Table 12: NDCG@10 on the R2MED benchmark. All models rerank the top-100 documents retrieved by ReasonEmbed (4B) (reason-embed-qwen3-4b-0928).

Relevance Evaluation Prompt Template

You are asked to evaluate the relevance between a **query** and **10 documents** based on the following **relevance definition**:
relevance_definition

Please follow the steps below carefully. Your reasoning should be clear, structured, and concise.

1. Query Analysis: analyze the query and describe what information would be most helpful for answering it.

- Focus on the key information needs implied by the query.
- Consider context, intent, and any specific constraints.
- Token limit: Maximum 256 tokens.

Output format:

[query_analysis]

2. Document Analysis: analyze each document individually. For each document:

- Discuss whether the document provides the key information required to answer the query.
- Identify any missing, incomplete, or irrelevant information.
- Provide structured reasoning (bullet points recommended).
- Token limit: Maximum 128 tokens per document.

Output format:

[1]: [doc_1_analysis]

[2]: [doc_2_analysis]

[3]: [doc_3_analysis] ...(repeat for all documents)

3. Relevance Annotation: for each document:

1. Evaluate each document independently to determine its **absolute relevance** to the query.
2. Assign an **integer relevance score (0–100)** according to the guide:
 - 81–100 (Highly Relevant): Fully and directly addresses the query; core, authoritative content.
 - 61–80 (Relevant): Mostly addresses the query; key info included, minor details may be missing.
 - 41–60 (Moderately Relevant): Partially addresses the query; on-topic but not comprehensive.
 - 21–40 (Slightly Relevant): Mentions query keywords; main topic is different, limited value.
 - 0–20 (Irrelevant): Does not address the query; off-topic.

Output format:

<scores>[s1] [s2] [s3] [s4] [s5] [s6] [s7] [s8] [s9] [s10]</scores>

Rules:

- Replace s1–s10 with actual integer scores
- Each score must be an integer between 0 and 100 and wrapped in []
- Scores must follow document order [1] to [10]
- Do NOT output the placeholder text or document ids
- Do NOT output explanations or extra text

4. Difficulty Annotation: rate the difficulty of assessing relevance on a 1–5 scale:

- 1 (Very Easy): Documents are clear and directly match the query.
- 2 (Easy): Most key info is present; minor reasoning needed.
- 3 (Medium): Partial info or multiple points require integration.
- 4 (Hard): Info is scattered, partially irrelevant, or ambiguous.
- 5 (Very Hard): Documents are complex, ambiguous, or noisy; deep reasoning required.

Output format:

<Difficulty>[1-5]</Difficulty>

<Justification>[1-2 sentence explanation]</Justification>

Input format:

Query: query_text

Documents: document_text

Table 13: Relevance Evaluation Prompt Template

Dataset	Relevance Definition
Biology	Given a query (biology post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
Earth Science	Given a query (earth science post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
Economics	Given a query (economics post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
Psychology	Given a query (psychology post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
Robotics	Given a query (robotics post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
StackOverflow	Given a query (Stack Overflow post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
Sustainable Living	Given a query (sustainable living post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.
LeetCode	Given a query (LeetCode problem) and a set of documents (coding problem solutions), the document is relevant to the query if the underlying algorithms or data structures used in the document can provide helpful insights for solving the problem in the query.
Pony	Given a query (Pony coding instruction) and a set of documents (Pony documentation passages), the document is relevant to the query if the Pony syntax described in the document is necessary for beginners with no prior knowledge of Pony to complete the coding instruction in the query.
AoPS	Given a query (math problem) and a set of documents (math problem solutions), the document is relevant to the query if the theorems used in the document can provide helpful insights for solving the problem in the query.
TheoremQA-Questions	Given a query (math problem) and a set of documents (math problem solutions), the document is relevant to the query if the theorems used in the document can provide helpful insights for solving the problem in the query.
TheoremQA-Theorems	Given a query (math problem) and a set of documents (math-related passages), the document is relevant to the query if the theorem described in the document can provide helpful insights for solving the problem in the query.

Table 14: Relevance definitions for the BRIGHT benchmark datasets.

Dataset	Relevance Definition
Biology	Given a query (biology question) and a set of documents (passages), the document is relevant to the query if it provides scientific explanations, biological mechanisms, or essential evidence that can help answer the query.
Bioinformatics	Given a query (bioinformatics question) and a set of documents (passages), the document is relevant to the query if it provides methods, databases, tools, or evidence that can help answer the query.
Medical-Sciences	Given a query (medical question) and a set of documents (passages), the document is relevant to the query if it provides scientific explanations, biological or clinical mechanisms, or evidence that can help answer the query.
MedXpertQA-Exam	Given a query (clinical diagnostic question) and a set of documents (passages), the document is relevant to the query if it provides clinical guidelines, evidence, reasoning, or explanations that can help answer the query.
MedQA-Diag	Given a query (clinical diagnostic question) and a set of documents (passages), the document is relevant if it provides evidence, clinical reasoning, or explanations that support the identification or confirmation of the correct diagnosis for the patient scenario described in the query.
PMC-Treatment	Given a query (clinical case and treatment question) and a set of documents (passages), the document is relevant if it provides evidence, guidelines, or explanations that support appropriate treatment strategies, drug choices, or therapeutic rationale for the patient case described in the query.
PMC-Clinical	Given a query (clinical case and diagnostic question) and a set of documents (passages), a document is considered relevant if it provides clinical evidence, case descriptions, or diagnostic reasoning that supports identifying the correct diagnosis for the patient described in the query.
Ilyi-Clinical	Given a query (clinical case) and a set of documents (passages), the document is relevant if it provides evidence, case descriptions, or diagnostic reasoning that supports identifying the correct diagnosis or understanding the patient’s clinical presentation.

Table 15: Relevance definitions for the R2MED benchmark datasets.

Dataset	Relevance Definition
TREC-COVID	Given a COVID-19 related query and a set of documents (passages), the document is relevant to the query if it answers the query.
DBPedia-Entity	Given a query and a set of documents (passages) describing an entity from DBpedia, the document is relevant to the query if the entity described in the document matches the query.
SciFact	Given a scientific claim query and a set of documents (passages), the document is relevant to the query if it provides evidence supporting or refuting the scientific claim.
NFCorpus	Given a question and a set of documents (passages), the document is relevant to the query if the document can best answer the question.
Signal	Given a news event or topic query and a set of documents (headlines or summaries), the document is relevant to the query if it reports on, summarizes, or directly relates to the same event or topic.
Robust04	Given an information-need query and a set of documents, the document is relevant to the query if it contains information that satisfies the intent or topic described in the query, even if phrased differently.
News	Given a query about a contemporary news topic or event and a set of documents (news article from The Washington Post), the document is relevant to the query if it discusses, explains, or provides factual coverage of that event or topic.

Table 16: Relevance definitions for the BEIR benchmark datasets.

Relevance Definition: Given a query (biology post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.

Query: I'm currently studying the cellular structure of bacteria and I've come across the fact that bacterial cells are typically surrounded by a cell membrane primarily composed of phospholipids. Unlike eukaryotic cells, bacteria usually lack large membrane-bound structures in their cytoplasm. Could someone elaborate on how the absence of these structures affects the functionality and organization of the bacterial cell? Specifically, how do bacteria manage to carry out complex processes without organelles like mitochondria or chloroplasts? Any insights or references would be greatly appreciated!

Output:

1. Query Analysis:

The query seeks explanations about bacterial cellular organization without membrane-bound organelles, specifically: - How the absence of organelles (e.g., mitochondria, chloroplasts) impacts functionality. - Mechanisms bacteria use for complex processes (e.g., energy generation, metabolism). - Insights into structural adaptations compensating for organelle lack. Domain experts require authoritative biological concepts about prokaryotic cell biology, membrane functions, and alternative intracellular structures.

2. Document Analysis:

- [1]: - Focuses on plant ATP synthesis in chloroplasts (eukaryotic). - Irrelevant to bacterial mechanisms; no bacterial context.
- [2]: - Describes phagocytosis in eukaryotes ('inside-out' membrane engulfment). - Mentions organelle origins but lacks bacterial adaptations.
- [3]: - Notes bacterial extracellular structures (cell wall, gas vacuoles). - Omits cytoplasmic organization or energy processes.
- [4]: - Confirms bacteria lack nuclei/membrane-bound organelles. - Does not explain functional adaptations.
- [5]: - States prokaryotes lack membrane-bound organelles and have circular DNA/ribosomes. - Superficial; misses bacterial-specific strategies.
- [6]: - Details nucleoid structure and ribosomal differences in bacteria. - Mentions intracellular storage granules but skips energy/metabolism mechanisms.
- [7]: - Compares mitochondrial/chloroplast membranes to prokaryotes. - Indirectly relevant to evolutionary origins but not functional adaptations.
- [8]: - Key info: Bacterial protein-bound organelles (e.g., carboxysomes), cytoskeleton, and membrane gradients for energy generation. - Directly addresses compensatory mechanisms.
- [9]: - Describes eukaryotic endomembrane systems (ER, Golgi). - Irrelevant to bacteria.
- [10]: - Key info: Bacterial plasma membrane folds for photosynthesis/electron transport; chlorosomes and carboxysomes as functional substitutes. - Directly explains organelle-free processes.

3. Relevance Annotation:

<score> [0] [20] [40] [65] [55] [70] [50] [90] [0] [85] </score>

- [1]: 0 (Irrelevant) – Eukaryotic photosynthesis. - [2]: 20 (Slightly Relevant) – Eukaryotic phagocytosis; tangential organelle mention. - [3]: 40 (Slightly Relevant) – Extracellular structures; misses core functionality. - [4]: 65 (Relevant) – Confirms organelle absence but lacks adaptations. - [5]: 55 (Moderately Relevant) – Basic prokaryote traits; no bacterial specifics. - [6]: 70 (Relevant) – Nucleoid/ribosome details; partial for organization. - [7]: 50 (Moderately Relevant) – Evolutionary clue; indirect functional insight. - [8]: 90 (Highly Relevant) – Protein-bound organelles, cytoskeleton, and membrane gradients for energy. - [9]: 0 (Irrelevant) – Eukaryotic endomembrane focus. - [10]: 85 (Highly Relevant) – Membrane folding, chlorosomes, carboxysomes for photosynthesis/metabolism.

4. Difficulty Annotation:

<Difficulty>3</Difficulty>

<Justification>Key documents (8,10) directly address bacterial adaptations, but scattered eukaryotic content (1,2,9) and partial/incomplete details (3,5,7) require careful filtering.</Justification>

Table 17: Case study of the Biology dataset in the BRIGHT benchmark.

Relevance Definition: Given a query (earth science post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.

Query: I'm currently researching atmospheric aerosols and their impact on climate, particularly over regions like China. I've come across some detailed studies on the optical, physical, chemical, and radiative properties of total columnar atmospheric aerosols. One of these studies involves the Sun-Sky Radiometer Observation Network (SONET) measurements. Could anyone provide insights on how SONET data contributes to our understanding of aerosol properties and climate effects over China? Additionally, how do these properties vary across different regions and what are the implications for global climate models?

Output:

1. Query Analysis:

The query seeks insights on: - SONET's role in understanding aerosol properties (optical, physical, chemical, radiative) and climate impacts over China. - Regional variations in aerosol properties across China. - Implications for global climate models. Domain experts require authoritative details on SONET's methodology, data applications, and empirical findings linking aerosols to climate effects. Context: Focus on China-specific studies with direct relevance to climate modeling.

2. Document Analysis:

[1]: SONET is used for classifying aerosol types in China (10 categories) using climate and socioeconomic factors. Discusses regional aerosol composition and polluted cloud verification. Missing explicit discussion of climate impacts or modeling implications.

[2]: Extends cloud-related analysis using OPAC software and dust transport links. Irrelevant to SONET aerosol property contributions.

[3]: Only references aerosol model clustering study. Incomplete; no SONET or climate linkage.

[4]: Focuses on polluted cloud transport via polarization algorithms. No SONET relevance.

[5]: Discusses polluted cloud identification and CALIPSO comparison. Missing SONET and climate relevance.

[6]: Repeats SONET-based aerosol classification and polluted cloud analysis. Partial relevance but lacks climate/model implications.

[7]: Same as [1], confirming SONET-based aerosol classification.

[8]: Mentions aerosol properties over China but lacks explicit SONET contribution or modeling insights.

[9]: Same as [1], SONET-based classification mentioned.

[10]: Uses MODIS/CALIPSO/CAMS for dust and irradiance modeling. Irrelevant to SONET.

3. Relevance Annotation:

<score> [65] [35] [25] [10] [30] [65] [65] [45] [65] [15] **</score>**

- [1], [6], [7], [9]: 65 (Relevant) — SONET used for aerosol classification in China, but limited climate/model discussion. - [2], [5]: 35, 30 (Slightly Relevant) — polluted cloud analysis, tangential to SONET focus. - [3]: 25 (Slightly Relevant) — general aerosol modeling mention, lacks detail. - [4]: 10 (Irrelevant) — cloud transport only. - [8]: 45 (Moderately Relevant) — aerosol properties mentioned but weak SONET linkage. - [10]: 15 (Irrelevant) — unrelated satellite/instrument focus.

4. Difficulty Annotation:

<Difficulty>4**</Difficulty>**

<Justification>Documents are fragmented and heavily focused on polluted cloud analysis rather than SONET's direct role in aerosol properties and climate modeling. Significant reasoning is required to extract sparse relevant evidence and filter noise from partially related content.**</Justification>**

Table 18: Case study of the Earth Science dataset in the BRIGHT benchmark.

Relevance Definition: Given a query (economics post) and a set of documents (passages), the document is relevant to the query if the critical concepts or theories discussed in the document can provide helpful insights for domain experts to answer the query.

Query: I'm exploring the relationship between exchange rates and economic growth, particularly how different studies have highlighted varying directions of causality and impact. For instance, some research suggests a unidirectional relationship where exchange rates significantly influence economic growth, as seen in Romania. Conversely, other studies indicate that economic growth can have a substantial effect on exchange rates, especially in certain South American and Asian countries. Additionally, there are findings that a high real exchange rate can stimulate economic growth, and undervaluation of the real exchange rate can positively affect growth, particularly in emerging economies. There is also evidence that the real exchange rate impacts growth indirectly through functional income distribution and technological innovation. Could anyone provide a detailed analysis or further studies that explore these relationships, especially in different economic contexts?

Output:

1. Query Analysis:

The query seeks studies on the bidirectional causality between exchange rates and economic growth, emphasizing: - Empirical evidence of unidirectional/directional causality (e.g., exchange rates → growth in Romania; growth → exchange rates in South America/Asia). - Mechanisms like real exchange rate undervaluation stimulating growth in emerging economies. - Indirect channels (e.g., income distribution, innovation). - Comparative analysis across economic contexts (developed vs. developing). Key needs: Causal direction studies, regional comparisons, mechanisms linking exchange rates to growth.

2. Document Analysis:

- [1]: Exchange rate depreciation reduces GDP in Liberia and appreciation harms growth in a 150-country study. Focuses on unidirectional effect (exchange rate → growth).
- [2]: Distinguishes developed vs developing economies: stability supports growth in developed countries, while depreciation may stimulate exports and growth in developing ones.
- [3]: Focuses on econometric testing (interest rate parity, stationarity) in Pakistan. No direct growth causality discussion.
- [4]: Exchange rate misalignment affects trade and GDP via net exports (e.g., 10% depreciation increases GDP by 1.5%).
- [5]: Same as [4], emphasizing trade-growth channel through exchange rate changes.
- [6]: Granger causality between inflation and exchange rates in Pakistan. No growth linkage.
- [7]: Inflation–exchange rate volatility across ASEAN and Latin America using GARCH models; limited growth discussion.
- [8]: Same as [4], focusing on exchange rate misalignment and trade impacts on GDP.
- [9]: Reference list only; no substantive content.
- [10]: Cointegration analysis of macro variables in Pakistan (interest rates, money supply). No growth–exchange rate focus.

3. Relevance Annotation:

<score> [65] [85] [15] [60] [60] [20] [35] [60] [5] [10] </score>

- [1]: 65 (Relevant) — exchange rate effects on growth (unidirectional evidence). - [2]: 85 (Highly Relevant) — strong regional comparison and mechanism differences. - [3]: 15 (Irrelevant) — econometric methodology only. - [4], [5], [8]: 60 (Moderately Relevant) — trade channel from exchange rate to GDP. - [6]: 20 (Irrelevant) — inflation focus, no growth link. - [7]: 35 (Slightly Relevant) — volatility analysis with weak growth connection. - [9]: 5 (Irrelevant) — references only. - [10]: 10 (Irrelevant) — macro modeling without target relationship.

4. Difficulty Annotation:

<Difficulty>4</Difficulty>

<Justification>Relevant information is scattered across documents; many contain technical noise or redundant content. Requires disentangling growth-specific insights from unrelated econometrics.**</Justification>**

Table 19: Case study of the Economics dataset in the BRIGHT benchmark.

Relevance Definition: Given a query (LeetCode problem) and a set of documents (coding problem solutions), the document is relevant to the query if the underlying algorithms or data structures used in the document can provide helpful insights for solving the problem in the query.

Query: Design a system that manages transactions for a database-like structure. Each transaction can either be explicit, where the user must manually commit or rollback, or implicit, where the system automatically commits if no errors occur. The system should support nested transactions, where a transaction can be started within another transaction. When an inner transaction is committed or rolled back, the outer transaction should remain unaffected unless all inner transactions are completed.

Implement a Database class with the following methods: - begin() - commit(transaction) - rollback(transaction) - transaction(implicit)

The system should support a context manager that automatically handles commit/rollback based on the implicit flag and exceptions, while correctly handling nested transactions.

Output:

1. Query Analysis:

The query requires designing a transaction system supporting: - Explicit transactions (manual commit/rollback) and implicit transactions (auto-commit if no errors). - Nested transactions where inner transactions don't affect outer ones until completion. - A 'Database class' with: - 'begin()': Starts a new transaction. - 'commit(transaction)'/rollback(transaction)': Targets specific transactions. - 'transaction(implicit)': Context manager handling auto-commit/rollback based on 'implicit' flag and exceptions. Key insights needed: 1. Transaction state tracking (nested hierarchy, active/committed/rolled back). 2. Context manager implementation for implicit/explicit control. 3. Isolation of inner/outer transactions during commits/rollbacks. 4. Error handling (exceptions triggering rollbacks). 5. Data structure for managing nested transactions (e.g., stack). Constraints: Nested transactions must remain isolated until all inner levels complete.

2. Document Analysis:

[1]: Implements transaction abort logic with state machine transitions (none, starting, committed, aborted). Missing nesting and context manager support.

[2]: Provides a basic transaction context manager with commit/rollback. Lacks nested transaction handling and implicit mode.

[3]: Uses subtransactions in DAO layer. Partially related but does not expose explicit transaction control interface.

[4]: Uses depth counter to support nested transactions and only commits at top level. Strongly relevant for nesting logic but lacks implicit transaction handling.

[5]: Implements global rollback mechanism without transaction isolation. Not suitable for nested design.

[6]: Provides commit with retry logic and metadata tracking. Missing rollback and nesting support.

[7]: Uses retry-based execution wrapper for transactions. Does not implement structured transaction hierarchy.

[8]: Uses stack-based transaction tracking with isolated rollback per level. Highly relevant for nested rollback logic but missing commit semantics.

[9]: Implements session-based context manager with commit/rollback on exception. Missing nested transaction control and explicit transaction targeting.

[10]: Explicitly forbids nested transactions. Contradicts problem requirements.

3. Relevance Annotation:

<score> [20] [40] [50] [75] [10] [20] [25] [65] [45] [5] </score>

- [1]: 20 (Irrelevant) — only basic state machine logic. - [2]: 40 (Slightly Relevant) — context manager exists but lacks nesting. -

[3]: 50 (Moderately Relevant) — partial subtransaction abstraction. - [4]: 75 (Highly Relevant) — depth-based nested transaction

handling. - [5]: 10 (Irrelevant) — no transaction isolation. - [6]: 20 (Irrelevant) — incomplete commit-only logic. - [7]: 25

(Slightly Relevant) — retry wrapper unrelated to structure. - [8]: 65 (Relevant) — stack-based nested rollback mechanism. - [9]:

45 (Moderately Relevant) — context manager but missing nesting. - [10]: 5 (Irrelevant) — contradicts nested requirement.

4. Difficulty Annotation:

<Difficulty>4</Difficulty>

<Justification>Documents are fragmented (e.g., [4] and [8] cover isolated aspects), require integration of nesting, context managers, and error handling, and some ([10]) contradict requirements. Ambiguity in transaction state tracking increases complexity.</Justification>

Table 20: Case study of the LeetCode dataset in the BRIGHT benchmark.

Relevance Definition: Given a query (Pony coding instruction) and a set of documents (Pony documentation passages), the document is relevant to the query if the Pony syntax described in the document is necessary for beginners with no prior knowledge of Pony to complete the coding instruction in the query.

Query: To implement a function in Pony that retrieves all flags for a multicast routing table (MRT) multicast forwarding cache (MFC) entry, define a primitive type `MrtMfcFlags` with functions returning bitwise flags in I32. The functions include `all()`, `border_vif()`, and `disable_wrongvif()`, each returning specific bit patterns using binary literals.

Output:

1. Query Analysis:

The query requires understanding of Pony syntax for defining primitive types using `primitive`, returning constant values with `fun` box methods, representing bitwise flags using binary literals (e.g., `0b00000001`), and encapsulating related flag constants within a single type.

2. Document Analysis:

- [1]: Defines flag primitives using bit shifting and integer constants. Demonstrates general flag encoding patterns but not MRT-specific usage.
- [2]: Introduces CPU feature flags with generic `Flags` abstraction. Shows parameterized flag structures but uses U64 instead of I32.
- [3]: Duplicate of [1], provides no additional information.
- [4]: Mentions MRT-related flag function returning I32 `val`. Partial but lacks implementation details.
- [5]: Direct MRT flag function naming with I32 return type. Closely aligned with query semantics but incomplete.
- [6]: Duplicate of [1], no new information.
- [7]: Explains generic flag system with storage type parameterization and feature flags (e.g., SSE, AVX). Useful for understanding structure but not MRT-specific.
- [8]: Describes flag operations such as `apply()`, `all()`, and `clear()`. Highly relevant for understanding flag manipulation logic.
- [9]: Defines a `Flag` interface requiring `value()` method and integer-based storage types. Important for understanding abstraction layer.
- [10]: Reiterates flag operations including `apply()` and `all()` methods. Reinforces core mechanisms.

3. Relevance Annotation:

`<score> [20] [30] [20] [65] [95] [20] [70] [75] [80] [60] </score>`

- [1], [3], [6]: 20 (Irrelevant) — generic flag encoding without MRT syntax relevance. - [2]: 30 (Slightly Relevant) — flag abstraction useful but type mismatch (U64). - [4]: 65 (Relevant) — MRT-related function with partial syntax coverage. - [5]: 95 (Highly Relevant) — direct match of MRT flag function structure. - [7]: 70 (Relevant) — strong conceptual support for flag system design. - [8]: 75 (Relevant) — core flag operations (`all()`, `apply()`). - [9]: 80 (Highly Relevant) — defines essential flag interface abstraction. - [10]: 60 (Moderately Relevant) — reinforces key operations but redundant.

4. Difficulty Annotation:

`<Difficulty>3</Difficulty>`

`<Justification>` Relevant information is distributed across multiple abstraction levels (primitive syntax, flag interfaces, and operation semantics). Although no single document fully explains MRT implementation, combining interface design and flag manipulation logic is sufficient for reconstruction.`</Justification>`

Table 21: Case study of the Pony dataset in the BRIGHT benchmark.

Relevance Definition: Given a query (math problem) and a set of documents (math problem solutions), the document is relevant to the query if the theorems used in the document can provide helpful insights for solving the problem in the query.

Query: A company sells a product and makes a profit of 25% on the cost price. If the company decides to increase the profit margin to 32%, the additional profit earned on each unit sold would be \$220. Considering the company operates in a competitive market and the increase in profit margin does not affect the sales volume, determine the cost price of each unit of the product. Furthermore, if the company sells 1,000 units of the product at the 32% profit margin, calculate the total additional profit earned compared to selling the same number of units at the original 25% profit margin.

Output:

1. Query Analysis:

The query requires applying the relationship between profit percentages on cost price. The key theorem is that the difference in profit percentages multiplied by the cost price equals the additional profit:

$$\text{Cost Price} = \frac{\text{Additional Profit}}{\text{Difference in Profit Rates}}$$

It also requires a direct scaling computation for total profit over multiple units:

$$\text{Total Additional Profit} = 220 \times 1000$$

2. Document Analysis:

- [1]: Focuses on deriving selling price from cost price and profit margin. Does not explicitly use profit-difference reasoning.
- [2]: Uses comparison between two profit scenarios to infer cost price. Partially relevant but includes extra variables such as purchase price changes.
- [3]: Directly applies profit difference theorem (e.g., $19\% - 12\% = 7\%$) to compute cost price from additional profit. Highly relevant and structurally identical to query requirement.
- [4]: Computes new profit percentage after price change. Not related to profit-difference theorem.
- [5]: Solves for selling price under different profit target. Does not use profit-difference reasoning.
- [6]: Similar to [2], introduces purchase price variation along with profit change. Partially relevant.
- [7]: Same structure as [5], focusing on target profit adjustment. Not relevant to theorem used in query.
- [8]: Same pattern as [2]/[6], involving extra purchase price assumptions. Partially relevant.
- [9]: Computes selling price for a given profit rate. Not relevant to profit difference structure.
- [10]: Similar to [2]/[6]/[8], includes additional purchase price variation. Partially relevant.

3. Relevance Annotation:

`<score> [20] [60] [95] [20] [40] [60] [20] [60] [20] [60] </score>`

[1]: Partial — explains profit-percentage relationships but not extra-profit comparison. [2]: Relevant — uses difference between two profit expressions to solve for an unknown. [3]: Complete — exact same theorem and reasoning structure as the query. [4]: Minimal — only recalculates profit percentage after price adjustment. [5]: Partial — involves changing profit percentages but not additional-profit logic. [6]: Relevant — comparative-profit algebra closely related to the query. [7]: Minimal — focuses on selling-price recalculation only. [8]: Relevant — uses profit-difference equations similar to the query. [9]: Slight — only generally related through percentage-profit manipulation. [10]: Relevant — strong similarity in comparing profit scenarios algebraically.

4. Difficulty Annotation:

`<Difficulty>3</Difficulty>`

`<Justification>Most documents share overlapping profit-related reasoning, but only one document directly applies the key theorem (profit difference method). The rest require filtering irrelevant variations such as purchase price changes or target profit adjustments.</Justification>`

Table 22: Case study of the TheoremQA Questions dataset in the BRIGHT benchmark.