R³A: Reinforced Reasoning for Relevance Assessment for RAG in User-Generated Content Platforms

Xiaowei Yuan[1,2], **Lei Jin**[3], **Haoxin Zhang**[3], **Ziyang Huang**[1,2],
Yan Gao[3], **Yi Wu**[3], **Yao Hu**[3], **Jun Zhao**[1,2], **Kang Liu**[1,2,*]
[1]The Key Laboratory of Cognition and Decision Intelligence for Complex Systems,
Institute of Automation, Chinese Academy of Sciences
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences
[3]Xiaohongshu Inc.

Abstract

Retrieval-augmented generation (RAG) plays a critical role in user-generated content (UGC) platforms, but its effectiveness critically depends on accurate query–document relevance assessment. Despite recent advances in applying large language models (LLMs) to relevance modeling, UGC platforms present unique challenges: 1) ambiguous user intent due to sparse user feedback in RAG scenarios, and 2) asymmetric relevance, where relevance is driven by localized answer-bearing content rather than global query–document similarity. To address these issues, we propose the Reinforced Reasoning model for Relevance Assessment (R³A), which decomposes relevance assessment into intent inference and evidence grounding. R³A leverages auxiliary high-clicked documents to infer latent query intent, and extracts verbatim evidence fragments to ground relevance decisions, reducing noise sensitivity and improving asymmetric relevance modeling. Experimental results demonstrate that R³A substantially outperforms strong baselines on offline benchmarks, while the distilled R³A-1.5B model achieves significant gains in large-scale online A/B testing, effectively balancing performance and practical deployability.

1 Introduction

Retrieval-augmented generation (RAG) systems have emerged as a critical paradigm in modern information retrieval, facilitating the generation of responses grounded in externally retrieved knowledge (Ram et al., 2023; Gao et al., 2024). Within large-scale user-generated content (UGC) platforms—encompassing product reviews, travelogues, and lifestyle narratives—RAG serves a critical function in search architectures. By integrating retrieval and generation capabilities, these systems can efficiently search over billions of user-contributed documents and produce concise, informative responses tailored to user queries (Li et al., 2025b; Zhang et al., 2025).

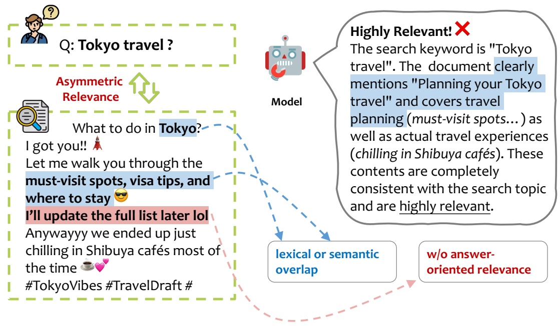


Figure 1: An illustrative example of erroneous relevance assessment for a query–document pair. The model is misled by superficial similarity in the noisy document, thereby failing to identify the absence of answer-bearing evidence required under asymmetric relevance.

A critical component of the RAG system is the relevance assessment module. This module quantitatively evaluates the semantic relevance between user queries and retrieved documents (Thomas et al., 2024), ensuring that the generation is accurately grounded in the query-related context. Recent advances in large language models (LLMs) have opened up new possibilities for relevance modeling by enabling fine-grained understanding of semantics and intent (Faggioli et al., 2023; Zheng et al., 2023), relying on prompting (MacAvaney and Soldaini, 2023; Upadhyay et al., 2024b) or supervised fine-tuning (SFT) methods (Zan et al., 2023; Fitte-Rey et al., 2025).

However, applying such models to UGC platforms introduces unique challenges. First, intent inference is constrained by the absence of click-through logs (Salemi and Zamani, 2024; Rathee et al., 2025), which conventional retrieval systems leverage to learn relevance–behavior mappings (Jiang et al., 2024). UGC-based RAG sys-

*Corresponding author.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track), pages 127–140

tems are typically supervised only at the response level, lacking the document-level interaction feedback necessary for effective **intent disambiguation**. In this context, intent represents the latent information need driving a query, which, on lifestyle-sharing UGC platforms, often transcends factual retrieval to encompass preferences for specific perspectives. As illustrated in Figure 1, a query such as "*Tokyo travel*" often implies a need for comprehensive guidance rather than isolated facts. Second, RAG systems must reason over **asymmetric relevance**, where document usefulness depends on the presence of answer-bearing fragments rather than global query–document similarity. This issue is amplified in UGC due to its informal and noisy nature, which obscures semantic cues and leads to relevance misestimation. As shown in Figure 1, the model is misled by superficial semantic overlap (e.g., "*must-visit spots*") despite the absence of substantive, answer-oriented content.

To address the above challenges, we propose a **R**einforced **R**easoning Model for **R**elevance **A**ssessment (R³A), which performs decomposed reasoning based on reinforcement learning (RL) algorithm. We argue that generating high-quality relevance assessment in UGC scenarios requires strong reasoning capabilities to address the challenges of ambiguous query intent and asymmetric relevance with noisy document. To better capture the user intent, the model input is augmented with a set of auxiliary in-platform high-clicked documents retrieved using the same query. These additional documents provide contexts to help the model infer the user's likely intent beyond the surface form of the query. Furthermore, the model is required to extract the most relevant answer fragment from the candidate document and determine relevance on the basis of this fragment. By compelling the model to isolate and justify its assessment using the fragments, this approach not only mitigates the impact of noise but also facilitates the evaluation of asymmetric relevance, determining the presence or absence of answer-bearing content that satisfies the user's information need.

Empirical results demonstrate that R³A consistently outperforms all baseline models in relevance assessment on both the industrial dataset NoteRel and the public benchmark T²Ranking (Xie et al., 2023). Furthermore, we distilled a compact R³A-1.5B model from the reinforced R³A-7B for online deployment. Notably, the distilled model not only exceeds the performance of its larger counterpart but also achieves significant gains over competing methods in large-scale online A/B testing.

The contributions of this paper are as follows:

- To tackle the unique challenges on the UGC platform, this paper proposes R³A method that performs decomposed reasoning over both ambiguous query and the asymmetric relevance with noisy document. It enhances reasoning for query disambiguation and evidence localization.
- On offline experiments, R³A consistently surpasses strong baselines. The distilled R³A-1.5B model outperforms its larger counterpart in online A/B testing, demonstrating the practical effectiveness of proposed R³A method.

2 Decomposed Reasoning for Relevance Assessment

This paper proposes the **R**einforced **R**easoning Model for **R**elevance **A**ssessment (R³A) method for UGC platforms, which enhances the reasoning capabilities of relevance modeling.

The overall framework of R³A is illustrated in Figure 2. After a cold-start initialization, the RL training procedure involves a two-stage interaction between the model and the environment (in-platform documents). *In the first stage*, a set of auxiliary in-platform documents $d'(\leq 4)$ is provided alongside the user query q to support the model in inferring latent query intent. *In the second stage*, the model extracts answer-bearing fragment from d to model the asymmetric relevance and ground its assessment in semantically aligned content.

For online deployment, we distill label logits from a reinforcement-trained large R³A model into a smaller model, which directly outputs relevance scores without generating reasoning traces.

2.1 Cold Start

Following prior work (Guo et al., 2025; Wei et al., 2025; Chen et al., 2025b), we first perform a cold-start phase using a 50k unlabeled dataset (see Section 3.1). This cold-start training on the relevance assessment task is designed to instill structured reasoning behavior and the desired output format, thereby improving training stability before RL. The structured chain-of-thought reasoning outputs are generated by DeepSeek-R1 (Guo et al., 2025)[1].

[1]Detailed prompt is provided in Appendix B.

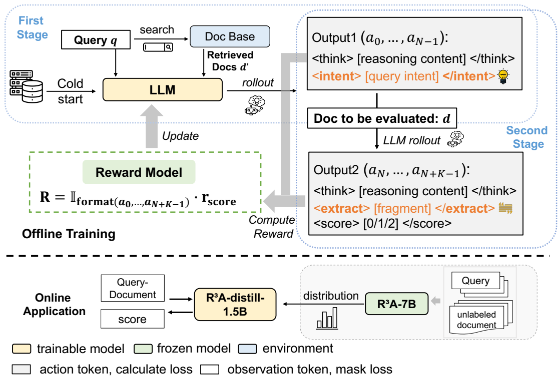


Figure 2: The overview of the R³A framework.

2.2 Decomposed Reasoning with RL

After the cold-start initialization, R³A is trained via the GRPO (Shao et al., 2024) algorithm. By estimating advantages through group-relative rewards instead of a dedicated critic model (as in PPO (Schulman et al., 2017)), GRPO facilitates more efficient large-scale training. We use GRPO primarily as a high-performance realization tool to showcase the improvements brought by our reasoning framework.

GRPO performs multiple rollouts per input and calculate the relative reward r within the group as the advantage A. It optimizes the following objective:

$$\mathcal{J}(\theta) = \mathbb{E}_{(q,d)\sim P(Q,D),\{\tau_i\}_{i=1}^{G}\sim\pi_{\theta_{old}}}\Bigg[\frac{1}{|G|}\sum_{i=1}^{|G|} \Bigg(\frac{1}{|\tau_i|}\sum_{t=1}^{|\tau_i|}\min\Big(r_\theta A_i, \text{clip}\big(r_\theta, 1-\epsilon, 1+\epsilon\big)A_i\Big)\Bigg)\Bigg],$$

$$r_\theta = \frac{\pi_\theta(a_{i,t}|x, d, a_{i,<t})}{\pi_{\theta_{old}}(a_{i,t}|x, d, a_{i,<t})} \tag{1}$$

where q and d denote the query and associated document sampled from the training distribution. Given an input (q, d) pair[2], a group G of trajectories τ_i is generated using the old policy $\pi_{\theta_{\text{old}}}$. Each trajectory τ_i comprises a sequence of actions a_i, representing the output of model reasoning. ϵ is the clipping ratio.

The term $A_i = \frac{r_i - \mu_r}{\sigma_r}$ represents the standardized advantage of trajectory τ_i, where r_i is the reward assigned to the trajectory, and μ_r and σ_r are the mean and standard deviation of the rewards within group G.

Reasoning on Query Intent. In the initial stage, the model interacts with a set of auxiliary, highly-clicked documents $d'(|d'| \leq 4)$ retrieved using the original query q from the platform. These documents supply additional contextual signals that enable the model to infer the user's latent intent beyond the query's surface form.

So the trajectory in the first stage can be represented as: $\tau_{i_1} = (q, d', a_{i,0}, a_{i,1}, \ldots, a_{i,N-1})$, where N denotes the number of action tokens. The action is structured token sequence using special tags, as pioneered by recent works (Li et al., 2025a; Song et al., 2025; Huang et al., 2025). Each action must contain a reasoning step (<THINK>) and an intent step (<INTENT>). The detailed system template is shown in Appendix C.1.

Reasoning with Asymmetric Relevance. As asymmetric relevance depends on the presence of answer-bearing segments rather than overall query–document similarity, the model is tasked with extracting verbatim, query-relevant fragments from candidate documents or returning NONE if no match exists. This formulation compels the model to ground its assessments in specific textual evidence, thereby ensuring directional, answer-centric judgments while mitigating the impact of noise.

In the extraction process, we enforce correctness strictly at the structural level—ensuring verbatim consistency with the document fragments—owing to the absence of ground-truth evidence annotations. The semantic accuracy and answer-oriented alignment of these extractions are intended to emerge through reinforcement-based reasoning training, which leverages available relevance labels to guide the model's optimization.

Thus, the complete trajectory in a rollout (two-stage interaction) can be represented as $\tau_i = (\tau_{i_1}, d, a_{N,i},\ a_{N+1,i}, \ldots, a_{N+K-1,i}, r)$, where K denotes the number of tokens and r denotes the reward to be calculated. Each action must contain a reasoning step (<THINK>), an extraction step (<EXTRACT>) and a final answer (<SCORE>). The detailed system template is shown in Appendix C.2.

Reward Function. We design the rule-based reward function such that a reward is granted if and only if the LLM-generated output fully conforms to all specified reasoning and answer formats, as well as the extraction consistency requirement. The total reward R is defined as:

$$R = \mathbb{I}_{\text{format}} \cdot r_{score}, \quad r_{\text{score}} = \begin{cases} 1 & s_{pred} == s_{gold} \\ \lambda & |s_{pred}\text{-}s_{gold}| = 1 \\ 0 & |s_{pred}\text{-}s_{gold}| = 2 \end{cases} \tag{2}$$

[2]The retrieved document d', based on the input query and document d, is omitted here for brevity.

where $\mathbb{I}_{\text{format}}$ is an indicator function that equals 1 if the trajectory format is correct, and 0 otherwise. The r_{score} measures the correctness of the model's prediction s_{pred} compared to the gold score s_{gold}. The hyperparameter $\lambda \in [0, 1)$ is introduced to impose a soft penalty. The value of λ balances the trade-off between strict correctness and leniency in reward shaping. In Exp. 3.4, we investigate the impact of varying λ on model performance.

2.3 Distillation

For online deployment, although RL with explicit reasoning can substantially improve performance, it also incurs prohibitive inference costs due to the generation of long reasoning traces. To reconcile these objectives, we deploy a distillation scheme that preserves the teacher's performance capabilities while eliminating the runtime overhead of trace generation. Concretely, we use a large reinforced R³A teacher (R³A-7B) to distill into a compact 1.5B student model on 50k samples. The distilled 1.5B model takes the query-document pair as input and directly outputs the relevance score. Unlike the teacher model, it does not generate reasoning traces, which minimizes online latency.

Formally, let $p_S(\cdot \mid x; \theta_S)$ denote the student's predictive distribution obtained by applying a softmax to its logits, and let $p_T(\cdot \mid x; \theta_T)$ be the final score distribution from the teacher model. The training objective is:

$$\mathcal{L} = -\mathbb{E}_{x\sim\mathcal{D}}\left[\sum_k p_T(k \mid x; \theta_T) \log p_S(k \mid x; \theta_S)\right] \quad (3)$$

3 Experiment

In this section, we conduct both offline and online experiments to evaluate the performance of the proposed R³A method.

3.1 Datasets

We evaluate relevance assessment performance on both an internal industrial dataset, **NoteRel**, and a public benchmark, **T²Ranking**. Summary statistics for the two datasets are reported in Table 1.

Datasets	NoteRel		T²Ranking	
	Training	Test	Training	Test
#0	2002	300	\	1000
#1	1858	300	\	1000
#2	1999	300	\	1000

Table 1: Summary statistics of datasets.

NoteRel. To assess relevance in a practical setting, we constructed NoteRel, a dataset derived from our online RAG system deployed on Xiaohongshu.

We extracted real-world user queries and their corresponding retrieved documents from online system logs. To generate high-quality training samples, we utilized document citation signals as a preliminary filter. For each query and candidate documents, the response generator performed M forward passes. A document was classified as a high-confidence positive sample if it was cited in at least N generated responses; otherwise, it was treated as a challenging hard negative. In our implementation, we set $M = 5$ and $N = 2$. To further account for distributional noise, we supplemented the dataset with random negatives sampled from the global corpus.

To establish "gold" standard labels, we collected 7k samples for professional human annotation across three categories: **0-Irrelevant, 1-Partially Relevant, and 2-Highly Relevant**. A detailed description of the dataset and data preparation procedures is provided in Appendix D. Additionally, 50k unlabeled documents were collected to support cold-start training and distillation phases.

T²Ranking. To ensure the generalizability of our framework, we also utilized T²Ranking, a large-scale retrieval benchmark comprising over 300k real-world queries and web passages. For our experiments, we sampled 3k records from the test set. T²Ranking provides fine-grained relevance labels on a 4-level scale (0–3). To maintain consistency with the NoteRel schema, we aligned the 2-level and 3-level labels in T²Ranking with the Highly Relevant (2) class in our industrial dataset.

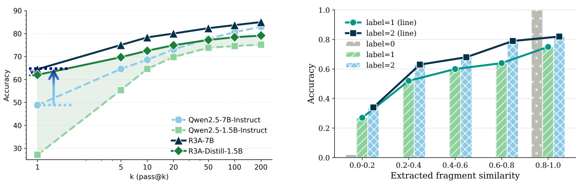
3.2 Settings

Baselines. To demonstrate the effectiveness of R³A, we compare it with the following baselines[3]: (1) **Prompting (UMbrela)** (Upadhyay et al., 2024b): It is a prompting-based method that provides a step-by-step guide to structure the relevance labeling task, thereby facilitating more nuanced reasoning by the LLM. In our experiment, we reproduce the UMbrela method using several LLMs, including QwQ (Team Qwen, 2024), DeepSeek-R1 (Guo et al., 2025), GPT-4o (Hurst et al., 2024), among others. (2) **SFT** (Fitte-Rey et al., 2025): This method introduces a framework that directly fine-tunes the model using relevance label. To en-

[3]All methods are conducted in a zero-shot manner.

Method	NoteRel						T²Ranking					
	F1-Score			AUC		Acc.	F1-Score			AUC		Acc.
	0	1	2	0/12	01/2		0	1	2	0/12	01/2	
Prompting												
UMbrela w/o parameter update												
Qwen2.5-7B	48.1	46.6	54.6	65.1	66.3	49.6	47.7	30.1	60.0	65.2	65.9	48.9
QwQ-32B	62.3	43.6	63.5	70.3	69.4	55.9	48.2	33.8	**67.1**	65.5	66.9	54.0
DeepSeek-V3	56.6	46.3	60.4	69.4	69.8	54.1	49.8	32.7	65.8	66.5	67.6	52.3
DeepSeek-R1	61.9	42.8	63.0	72.1	71.8	56.0	46.5	31.9	66.9	64.3	63.2	53.6
GPT-4o	63.3	53.8	59.6	73.0	69.8	58.2	50.9	33.4	66.2	66.3	67.1	**54.2**
Supervised Fine-Tuning												
w/ parameter update												
Qwen2.5-1.5B	68.4	51.8	57.7	76.3	68.5	59.0	48.3	36.0	50.4	67.6	62.9	46.3
+ pretrained	70.1	55.3	56.0	77.3	68.2	60.0	50.7	33.3	58.2	69.1	65.4	48.3
Qwen2.5-7B	71.0	55.6	52.1	78.2	66.8	60.3	49.7	32.2	62.8	66.4	67.5	50.7
+ pretrained	69.9	54.9	53.9	77.0	67.6	61.4	51.2	31.6	65.3	66.9	67.5	51.4
Reinforcement Learning												
w/ parameter update on Qwen2.5-1.5B												
R1-Zero	66.7	45.3	58.5	75.1	68.2	56.8	47.9	26.8	62.3	65.0	64.1	46.2
R1	70.7	46.9	62.0	78.3	69.0	59.7	51.1	25.4	63.6	65.6	64.8	47.7
R³A-Zero	71.3	49.8	60.5	79.4	68.3	60.4	48.0	24.8	61.8	64.4	65.0	47.9
R³A	72.0	51.5	62.9	80.8	69.6	61.7	49.8	30.3	62.7	66.1	66.4	49.0
w/ parameter update on Qwen2.5-7B												
R1-Zero	67.5	47.2	62.5	75.6	71.7	59.3	48.0	24.8	61.8	64.4	65.0	49.9
R1	74.4	49.6	63.7	81.2	72.7	63.2	50.8	27.3	57.7	66.4	65.6	51.4
R³A-Zero	75.8	51.7	63.3	82.4	72.5	63.6	50.8	33.2	63.5	67.3	64.8	51.2
R³A	**77.1**	**56.0**	**64.2**	**83.1**	**73.3**	**65.2**	**52.1**	**34.3**	66.6	**67.5**	**67.8**	53.9
Distilling (Online Serving)												
R³A-Distill-1.5B	71.4	55.9	60.3	78.3	70.5	62.0	50.4	29.8	64.9	66.3	67.2	51.3

Table 2: Overall performance on the test set of NoteRel and T²Ranking datasets. The labels 0, 1, and 2 indicate "Irrelevant", "Partially Relevant", and "Highly Relevant", respectively.



(a) Pass@k performance comparison across different models

(b) Correlation between extracted fragment similarity and accuracy

Figure 3: Performance analysis of R³A model.

sure a fair comparison, we also pre-train the model using cold-start data before fine-tuning. (3) **R1-Zero/R1** (Guo et al., 2025): This method employs rule-based GRPO to encourage the model to engage in explicit reasoning during relevance assessment. The model's output format is constrained to the <THINK> and <SCORE> tags. The R1-Zero refers to RL initiated without the cold-start strategy. (4) **Distillation**: The distilled model is trained on a 1.5B backbone using the same SFT approach. It relies solely on score labels produced by the R³A-7B model on 50k cold-start data. We deploy this version in our online system for better inference speed and overall throughput.

Models. For SFT and RL-based methods, we explore LLMs using instruction-tuned Qwen2.5 models (Yang et al., 2024) ranging from 1.5B to 7B parameters. All models are trained on the training split of the NoteRel dataset, with detailed training

Time / 8 A100 GPU (in seconds)	Number of Evaluated Documents				
	1	10	100	1,000	10,000
R³A-7B	0.6385	1.1718	3.3774	43.7740	493.7861
R³A-Distill-1.5B	0.1396	0.1717	0.3332	2.3673	10.9421

Table 3: Inference latency comparison of R³A-7B and R³A-Distill-1.5B models.

configurations provided in Appendix E and training logs presented in Appendix F.

Metrics. We use F1 score (macro-averaged), Accuracy, and one-vs-rest AUC metrics ($\text{AUC}_{0/12}$ and $\text{AUC}_{01/2}$) following previous works (Welleck, 2016; Chen et al., 2024).

3.3 Overall Performance

As presented in Table 2, our method R³A consistently achieves superior performance over all baseline approaches on the NoteRel test set. Moreover, on the out-of-distribution benchmark T²Ranking, R³A achieves performance comparable to GPT-4o, indicating that the proposed approach generalizes effectively to datasets from other commercial domains. We present several cases in Appendix G.

Notably, the R³A-Zero models, trained without cold-start initialization, already surpass both R1-Zero and SFT baselines, underscoring the efficacy of R³A in exploiting reasoning signals to improve performance even under constrained initialization settings. When full RL training is applied, R³A further extends its advantage, yielding state-of-the-art results across most evaluation metrics.

Superior Distillation. Remarkably, the distilled model R³A-Distill-1.5B not only retains the performance gains of its 1.5B RL-trained counterpart, but also outperforms the larger 7B SFT model. These results indicate that the knowledge distilled from R³A effectively preserves essential relevance assessment capabilities in a smaller model.

As shown in Table 3, we compare inference time before and after model distillation on NoteRel dataset. The results indicate that the R³A-7B model's inference latency increases almost linearly with the number of evaluated documents. In contrast, distillation substantially reduces inference cost by eliminating the need to output intermediate reasoning steps, thereby mitigating latency and improving efficiency.

Performance on Pass@k Figure 3a shows that while base instruction-tuned models improve with increasing k, their performance remains substantially below that of R³A-trained models, particu-

Method	F1-Score			AUC		Accuracy
	0	1	2	0/12	01/2	
R³A-7B ($\lambda = 0$)	**77.1**	**56.0**	**64.2**	**83.1**	**73.3**	**65.2**
Format Variants						
w/o `<intent>`	76.0	55.9	63.7	82.6	72.8	64.7
w/o `<extraction>`	74.9	52.7	63.2	81.1	72.4	63.9
*Input Variant**						
`w/o retrieval`	75.7	54.4	63.9	82.4	72.9	64.6
`w/ half noisy docs`	76.3	55.5	63.3	82.7	72.6	64.8
`w/ all noisy docs`	75.4	54.2	63.3	81.9	71.8	64.3
Reward Variants						
$\lambda = 0.5$	75.3	52.5	59.7	82.0	69.9	62.0
$\lambda = 0.2$	74.9	53.6	62.9	81.9	72.3	62.5
$\lambda = 0.1$	72.6	48.4	60.3	80.2	70.8	61.4
Interaction Stage Variant						
single stage †	73.8	49.1	58.3	80.7	68.5	60.6

* "Without retrieval" denotes the removal of retrieved in-platform documents. "half/all noisy docs" denote that half or all of the retrieved documents are randomly replaced with irrelevant ones.

† Both the retrieved documents and the document to be evaluated are input into the model, which is required to analyze the intent, extract the fragment and assess the relevance.

Table 4: Ablation study on the R³A method on NoteRel dataset.

larly at low k (e.g., pass@1 and pass@5), where robust reasoning is critical. Consistent with prior findings (Deng et al., 2025), RL yields pronounced gains in low-sampling regimes, enabling R³A-7B to consistently outperform all baselines across pass@k. Notably, the distilled R³A-Distill-1.5B model closely approaches its 7B teacher and achieves over 10% absolute improvement over Qwen2.5-1.5B-Instruct at pass@1, demonstrating effective knowledge transfer and strong deployability under efficiency constraints.

Role of Extraction in Reasoning Figure 3b analyzes the relationship between the similarity of reasoning fragments extracted by R³A-7B and those generated by GPT-4o under identical prompts, and its impact on relevance prediction accuracy. The results show a strong positive correlation, with near-optimal performance at high similarity levels (0.8–1.0) and sharp degradation at low similarity (0.0–0.2), particularly for partially relevant cases (score = 1). These findings highlight fragment extraction as a central mechanism in R³A, enabling accurate asymmetric relevance assessment while reducing the influence of document noise.

3.4 Ablation Study

Results in Table 4 validate each component's contribution. *Format Variants* reveal that omitting extraction reasoning causes the most significant degradation, emphasizing the necessity of document-based grounding. Accurate *retrieval in input* enhances intent inference by supplying informative contextual

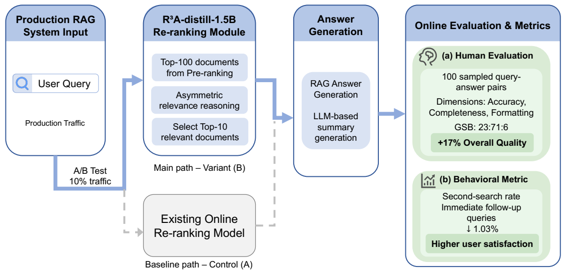


Figure 4: Overview of the online deployment and evaluation pipeline of R³A-Distill-1.5B.

cues. Regarding *Reward Variants*, $\lambda = 0$ is optimal for maintaining strict class differentiation. Lastly, collapsing the Two-stage Interaction into a single stage reduces accuracy to 60.6%, suggesting that excessive context length diminishes the model's focus on the document to be evaluated.

3.5 Online Performance

As illustrated in Figure 4, we deploy *R³A-Distill-1.5B* as the re-ranking module in the RAG system of Xiaohongshu and evaluate it via a one-week online A/B test on 10% of live traffic. The model re-ranks the top 100 pre-retrieved documents and selects the top 10 for answer generation.

We first perform a human evaluation using 100 randomly sampled query–answer pairs from the system logs. Compared with our existing online model (Fitte-Rey et al., 2025), the distilled model achieves a Good:Same:Bad (GSB) distribution of $23 : 71 : 6$, corresponding to a 17% improvement in overall answer quality. Moreover, using second-search rate as a proxy for user satisfaction, the distilled model achieves a 1.03% reduction, indicating improved effectiveness and reduced need for query reformulation.

By eliminating runtime reasoning trace generation, the distilled model achieves an average response time (RT) of approximately 150ms. Each single GPU in the production cluster handles a throughput of 200 QPS (Queries Per Second). To support the massive live traffic of the platform, we have deployed the distilled model on a cluster of 1,000 GPUs.

4 Related Work

Relevance modeling evaluates the extent to which a document satisfies a user query. Traditional human-annotated approaches are costly and prone to subjectivity, prompting interest in the "LLM-as-a-Judge" paradigm (Zheng et al., 2023).

Faggioli et al. (2023) are among the first to investigate a range of human–machine collaboration strategies in which LLMs assist in relevance judgment. Building upon this, automated evaluations using LLMs have combined various prompting techniques such as zero-shot, one-shot (MacAvaney and Soldaini, 2023), or few-shot learning (Thomas et al., 2024; Upadhyay et al., 2024a,b).

Another line of work (Ma et al., 2024; Fitte-Rey et al., 2025) involves training dedicated LLMs for assessment tasks. Both Ma et al. (2024) and Fitte-Rey et al. (2025) explore the use of smaller, fine-tune LLMs for relevance assessment in real-world scenarios. However, such fine-tuned judge models often function as task-specific classifiers, thereby inheriting certain limitations in generalizability and reasoning capacity (Huang et al., 2024). More recently, models such as JudgeLRM (Chen et al., 2025a) and Rank-R1 (Zhuang et al., 2025) have emerged, explicitly incorporating reasoning across different assessment tasks through RL with outcome-driven rewards.

It is worth noting that some recent approaches have also integrated user behavior signals into the model input or reward to tackle personalized relevance. For example, Chen et al. (2024) proposed the ProPBP that feeds user interaction data (clicks, dwell time, etc.) into the LLM judge to align its decisions with individual user preferences. However, such behavior signals are largely unavailable in the RAG system, rendering these methods inapplicable.

5 Conclusion

This paper proposes the R³A method, a novel decomposed reasoning framework tailored for relevance assessment for RAG system in UGC scenarios. Empirically, R³A exhibits strong capabilities in relevance assessment task in UGC scenarios.

Limitations

Despite the effectiveness of R³A in enhancing relevance assessment for RAG systems within UGC scenarios, several limitations remain. R³A is primarily trained on an industry-specific UGC dataset. As demonstrated in Exp.3.3, our method consistently outperforms all baseline approaches on the NoteRel test set. On the out-of-distribution benchmark T²Ranking, the performance gap of R³A is notably smaller than that observed on NoteRel. During in-domain training, the model learns to reason about user query intent within UGC communities. However, UGC platforms and general-purpose search engines exhibit distinct contextual characteristics. When the model applies reasoning patterns learned from UGC data to assess relevance between queries and web passages in non-UGC settings, the contextual mismatch limits the magnitude of performance gains. Nevertheless, the proposed approach demonstrates reasonable generalization to datasets from other commercial domains and remains computationally efficient and well-suited for RAG applications in UGC scenarios.

Acknowledgements

This work was supported by Beijing Natural Science Foundation (L243006), the National Natural Science Foundation of China (No. U24A20335), the independent research project of the Key Laboratory of Cognition and Decision Intelligence for Complex Systems.

References

Nuo Chen, Zhiyuan Hu, Qingyun Zou, Jiaying Wu, Qian Wang, Bryan Hooi, and Bingsheng He. 2025a. Judgelrm: Large reasoning models as a judge. *CoRR*, abs/2504.00050.

Shuang Chen, Yue Guo, Zhaochen Su, Yafu Li, Yulun Wu, Jiacheng Chen, Jiayu Chen, Weijie Wang, Xiaoye Qu, and Yu Cheng. 2025b. Advancing multimodal reasoning: From optimized cold start to staged reinforcement learning. *CoRR*, abs/2506.04207.

Zeyuan Chen, Haiyan Wu, Kaixin Wu, Wei Chen, Mingjie Zhong, Jia Xu, Zhongyi Liu, and Wei Zhang. 2024. Towards boosting llms-driven relevance modeling with progressive retrieved behavior-augmented prompting. *arXiv preprint arXiv:2408.09439*.

Jia Deng, Jie Chen, Zhipeng Chen, Daixuan Cheng, Fei Bai, Beichen Zhang, Yinqian Min, Yanzipeng Gao, Wayne Xin Zhao, and Ji-Rong Wen. 2025. From trial-and-error to improvement: A systematic analysis of llm exploration mechanisms in rlvr. *Preprint*, arXiv:2508.07534.

Guglielmo Faggioli, Laura Dietz, Charles L. A. Clarke, Gianluca Demartini, Matthias Hagen, Claudia Hauff, Noriko Kando, Evangelos Kanoulas, Martin Potthast, Benno Stein, and Henning Wachsmuth. 2023. Perspectives on large language models for relevance judgment. In *Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2023, Taipei, Taiwan, 23 July 2023*, pages 39–50. ACM.

Quentin Fitte-Rey, Matyas Amrouche, and Romain Deveaud. 2025. Augmented relevance datasets with fine-tuned small llms. *CoRR*, abs/2504.09816.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. 2024. Retrieval-augmented generation for large language models: A survey. *Preprint*, arXiv:2312.10997.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, and 80 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *CoRR*, abs/2501.12948.

Jian Hu, Xibin Wu, Weixun Wang, Xianyu, Dehao Zhang, and Yu Cao. 2024. Openrlhf: An easy-to-use, scalable and high-performance RLHF framework. *CoRR*, abs/2405.11143.

Hui Huang, Yingqi Qu, Jing Liu, Muyun Yang, and Tiejun Zhao. 2024. An empirical study of llm-as-a-judge for LLM evaluation: Fine-tuned judge models are task-specific classifiers. *CoRR*, abs/2403.02839.

Ziyang Huang, Xiaowei Yuan, Yiming Ju, Jun Zhao, and Kang Liu. 2025. Reinforced internal-external knowledge synergistic reasoning for efficient adaptive search agent. *CoRR*, abs/2505.07596.

Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Madry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, and 79 others. 2024. Gpt-4o system card. *CoRR*, abs/2410.21276.

Guanying Jiang, Lingyong Yan, Haibo Shi, and Dawei Yin. 2024. The real, the better: Aligning large language models with online human behaviors. *Preprint*, arXiv:2405.00578.

Xiaoxi Li, Guanting Dong, Jiajie Jin, Yuyao Zhang, Yujia Zhou, Yutao Zhu, Peitian Zhang, and Zhicheng Dou. 2025a. Search-o1: Agentic search-enhanced large reasoning models. *CoRR*, abs/2501.05366.

Yuchen Li, Hengyi Cai, Rui Kong, Xinran Chen, Jiamin Chen, Jun Yang, Haojie Zhang, Jiayi Li, Jiayi Wu, Yiqun Chen, Changle Qu, Keyi Kong, Wenwen Ye, Lixin Su, Xinyu Ma, Long Xia, Daiting Shi, Jiashu Zhao, Haoyi Xiong, and 2 others. 2025b. Towards ai search paradigm. *Preprint*, arXiv:2506.17188.

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, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 2421–2425. ACM.

Sean MacAvaney and Luca Soldaini. 2023. One-shot labeling for automatic relevance estimation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2230–2235. ACM.

Ori Ram, Yoav Levine, Itay Dalmedigos, Dor Muhlgay, Amnon Shashua, Kevin Leyton-Brown, and Yoav Shoham. 2023. In-context retrieval-augmented language models. *Transactions of the Association for Computational Linguistics*, 11:1316–1331.

Mandeep Rathee, Venktesh V, Sean MacAvaney, and Avishek Anand. 2025. Test-time corpus feedback: From retrieval to RAG. *CoRR*, abs/2508.15437.

Alireza Salemi and Hamed Zamani. 2024. Evaluating retrieval quality in retrieval-augmented generation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 2395–2400. ACM.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *Preprint*, arXiv:1707.06347.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *CoRR*, abs/2402.03300.

Huatong Song, Jinhao Jiang, Yingqian Min, Jie Chen, Zhipeng Chen, Wayne Xin Zhao, Lei Fang, and Ji-Rong Wen. 2025. R1-searcher: Incentivizing the search capability in llms via reinforcement learning. *CoRR*, abs/2503.05592.

Team Qwen. 2024. Qwq: Reflect deeply on the boundaries of the unknown.

Paul Thomas, Seth Spielman, Nick Craswell, and Bhaskar Mitra. 2024. Large language models can accurately predict searcher preferences. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 1930–1940. ACM.

Shivani Upadhyay, Ehsan Kamalloo, and Jimmy Lin. 2024a. Llms can patch up missing relevance judgments in evaluation. *CoRR*, abs/2405.04727.

Shivani Upadhyay, Ronak Pradeep, Nandan Thakur, Nick Craswell, and Jimmy Lin. 2024b. UMBRELA: umbrela is the (open-source reproduction of the) bing relevance assessor. *CoRR*, abs/2406.06519.

Lai Wei, Yuting Li, Kaipeng Zheng, Chen Wang, Yue Wang, Linghe Kong, Lichao Sun, and Weiran Huang. 2025. Advancing multimodal reasoning via reinforcement learning with cold start. *CoRR*, abs/2505.22334.

Sean J. Welleck. 2016. Efficient AUC optimization for information ranking applications. In *Advances in Information Retrieval - 38th European Conference on*

IR Research, ECIR 2016, Padua, Italy, March 20-23, 2016. Proceedings, volume 9626 of *Lecture Notes in Computer Science*, pages 159–170. Springer.

Xiaohui Xie, Qian Dong, Bingning Wang, Feiyang Lv, Ting Yao, Weinan Gan, Zhijing Wu, Xiangsheng Li, Haitao Li, Yiqun Liu, and Jin Ma. 2023. T2ranking: A large-scale chinese benchmark for passage ranking. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2681–2690. ACM.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, and 22 others. 2024. Qwen2.5 technical report. *CoRR*, abs/2412.15115.

Wen Zan, Yaopeng Han, Xiaotian Jiang, Yao Xiao, Yang Yang, Dayao Chen, and Sheng Chen. 2023. Spm: Structured pretraining and matching architectures for relevance modeling in meituan search. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, page 4923–4929. Association for Computing Machinery.

Yu Zhang, Shutong Qiao, Jiaqi Zhang, Tzu-Heng Lin, Chen Gao, and Yong Li. 2025. A survey of large language model empowered agents for recommendation and search: Towards next-generation information retrieval. *Preprint*, arXiv:2503.05659.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging llm-as-a-judge with mt-bench and chatbot arena. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.

Shengyao Zhuang, Xueguang Ma, Bevan Koopman, Jimmy Lin, and Guido Zuccon. 2025. Rank-r1: Enhancing reasoning in llm-based document rerankers via reinforcement learning. *CoRR*, abs/2503.06034.

A LLM Usage Disclosure

We use LLM for paper writing to check grammar and boost the clarity. We do not use LLM to generate experiment code and analysis.

B UMbrela Prompt

The following presents the full prompt used in UMbrela method, which is also employed by DeepSeek-R1 to generate reasoning chains and answers on unlabeled data.

Prompt

System
You are a relevance assessor working on a user-generated content platform.
User
Given a query and a document, you must provide a score on an integer scale of 0 to 2 with the following meanings:
0 = represent that the document has nothing to do with the query
1 = represents that the document has some answer for the query, but the answer may be a bit unclear, or hidden amongst extraneous information
2 = represents that the document is dedicated to the query and contains the exact answer
Important Instruction:
Assign category 1 if document presents something very important related to the entire topic but also has some extra information and category 2 if the document only and entirely refers to the topic. If none of the above satisfies give it category 0.
Please determine the primary intent behind a user's search query, using both your internal knowledge and the provided context.

Your response must strictly follow the format:
`<think> [the reasoning content] </think>`
`<score> [0/1/2] </score>`
Input
[query]: $\{query\}$
[document to be evaluated]: $\{doc\}$
Assistant

C Instruction Template for R³A

C.1 Prompt in the First Stage

The following presents the full prompt used in the first stage interaction.

Prompt in the 1st Stage

System
You are a content understanding engineer working on a user-generated content platform.
User
Please determine the primary intent behind a user's search query, using both your internal knowledge and the provided context.
Your input consists of the [query] and the [in-platform documents] retrieved based on that query. The latter is intended to assist in judging the user's intent but may contain irrelevant content. The search query should be considered the primary reference. Please carefully analyze the given [query] and the corresponding [in-platform documents] to infer the underlying query intent.

Your response must strictly follow the format:
`<think> [the reasoning content] </think>`
`<intent> [inferred user intent] </intent>`
Input
[query]: $\{query\}$
[in-platform documents]: $\{docs\}$
Assistant

C.2 Prompt in the Second Stage

The following presents the full prompt used in the second stage interaction.

Prompt in the 2nd Stage

User
Please assess the relevance of the [document to be evaluated] based on the user's input [query] and the inferred [intent], and extract the relevant fragment of the document accordingly.
Scoring Criteria
0 = not relevant, the document has nothing to do with the query.
1 = partially relevant, the document is relevant to the query but partly answers it.
2 = highly relevant, the document is dedicated to the query and contains the exact answer.
Extraction Guidelines
1. Extract the content from the [document to be evaluated] that is strictly relevant to the query and can help answer the query. This may include paragraphs, sentences, or even individual phrases.
2. The extracted content must come directly from the original document, with all punctuation preserved.

Your response must strictly follow the format:
`<think> [the reasoning content] </think>`
`<extract> [fragment/none] </extract>`
`<score> [0/1/2] </score>`
Input
[document to be evaluated]: $\{doc\}$
Assistant

D Dataset Description

We collect a total of 50k unlabeled documents for cold-start training and distillation, and another 7k samples for human double-blind annotation. Our dataset has three classes:

- **0-Irrelevant**: There is a complete mismatch between the content of the query and the document.
- **1-Partially Relevant**: The document is relevant to the query and partly satisfies user's information needs.
- **2-Highly Relevant**: The document content is customized to satisfy the information needs of the query and precisely contains the answer to the query.

D.1 Data Strategy

Cold-Start (50k). Following the data collection strategy detailed in Sec 3.1, we utilized document citation signals as a preliminary filter to curate an approximately balanced set of 50k unlabeled query-document pairs. Subsequently, we employed DeepSeek-R1 to generate structured CoT reasoning outputs. This phase is designed to finstill structured reasoning behavior and the desired output format.

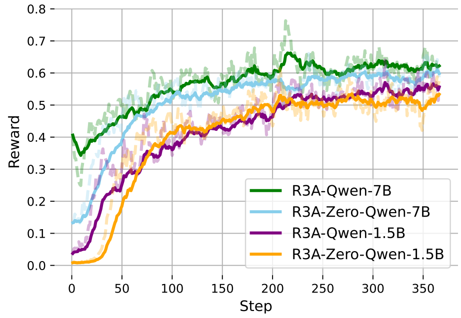
RL Training (NoteRel 6k). We utilize the NoteRel training set, comprising approximately 6k human-annotated samples as ground truth, to fine-tune all trainable models.

Distillation (50k). We employ the same 50k unlabeled query-document pairs from the cold-start phase. These samples are labeled by the R³A-7B model, providing both reasoning traces and final scores. The query-document pair inputs with only score outputs are then used to distill knowledge into a compact 1.5B student model.

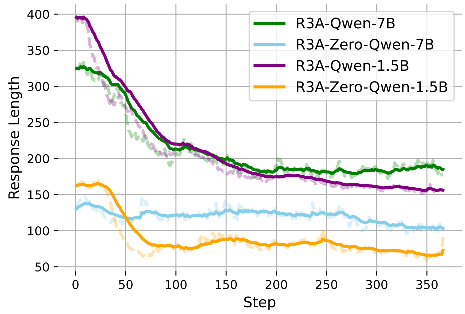
D.2 Human Annotation

To ensure the reliability of the annotations, we employed rigorous quality control measures, including:

1. Annotation Guidelines: Annotators were provided with detailed guidelines and training sessions to standardize their understanding of relevance criteria, minimizing potential biases in the annotations.
2. Inter-annotator Agreement: A subset of annotations was double-annotated by two independent annotators to assess the consistency of the judgments. Only those annotations with an agreement above a predefined threshold were retained for further use.



(a) Training Rewards



(b) Response Length

Figure 5: The training log of R³A-Zero-1.5/7B and R³A-1.5/7B, including the curve of training rewards and response length.

3. Feedback Loops: Regular feedback was provided to annotators to ensure continuous improvement of annotation quality during the process.

These measures were implemented to ensure that the annotations are reliable, consistent, and representative of the true relevance of query-document pairs.

E Implementation Details

We employ Qwen2.5-1.5B(-Instruct) and Qwen2.5-7B(-Instruct) as the initial models. Models with the "-Zero" suffix are trained without cold start from the Instruct model. We utilize the OpenRLHF (Hu et al., 2024) framework for training. GRPO (Shao et al., 2024) is used as the RL algorithm. We use the NoteRel as the training and test sets. We set the number of rollouts as 16 for one task. We set the learning rate as 5e-7, batch size as 32, training steps as 360. We set λ as 0 in reward function. We use 8 A100 GPUs for all the experiments.

F Training Log of R³A-7B/1.5B

Figure 5 shows that R³A models initialized with cold start exhibit faster reward growth and achieve higher final rewards compared to their R³A-Zero counterparts, underscoring the benefit of cold-start initialization. Although Zero models begin with lower rewards, they gradually acquire the desired format, demonstrating that decomposed reasoning remains effective even without prior initialization. In terms of response length, both R³A and R³A-Zero models rapidly converge to a stable and concise output length.

G Case Study

Representative outputs from the R^3A framework are provided in Tables 5 and 6. For brevity, the auxiliary high-clicked documents used for intent analysis are omitted to more clearly highlight the interaction between the query and document.

Example 1 illustrates a document with a gold score of 0, which was misclassified as partly relevant by the R1 model but correctly classified by R³A. The user query asked for the precise definition of an infinite series, whereas the document only contained related concepts from a table of contents and did not address the question directly. The R1 model was misled by superficial lexical matches and assigned a relevance score of 1, despite the document failing to meet the user's information need. In contrast, our model, R³A, successfully inferred the document's lack of relevance by identifying and reasoning over the actual content, ultimately producing the correct assessment. This case highlights a key strength of R³A: its ability to go beyond lexical similarity by integrating retrieved documents with deeper reasoning about content sufficiency, thereby producing more reliable relevance judgments in alignment with user intent.

In Example 2, the user inquires about the specific reasons for the absence of kindergarten tuition fees in Suzhou. The source document exhibits typical UGC noise, containing extraneous information such as personal anecdotes and traffic conditions. R³A demonstrates superior query disambiguation through its <INTENT> analysis, correctly identifying that the user's core intent is to verify the implementation details of a regional fee-waiver policy. By leveraging its decomposed reasoning capability, the model successfully isolates the pertinent policy evidence from the surrounding noise, achieving precise evidence localization (Score 2).

Example 3 highlights R³A's ability to handle partial informational alignment in complex scenarios. The query concerns general ticket-purchasing procedures for seniors, while the document provides instructions specific to a seasonal promotion in the first quarter of 2024. The <THINK> output analytically recognizes the asymmetric relationship between the general query and the time-constrained evidence. The model correctly identifies that the document provides valid procedural steps but is limited by its seasonal nature, leading to an accurate partial relevance score (Score 1). This demonstrates the framework's robustness in maintaining high reasoning precision even when evidence is bound by specific temporal or conditional constraints.

Example 1:
Query: What does infinite series mean?
Document: Science students at the University of Manchester, speed up! The most comprehensive calculus vocabulary list is here Hi dear students studying at UoM I'm sure your English is totally fine for daily communication, but when it comes to some technical terms, it can get tricky! Looking up unfamiliar words one by one is just too much trouble and really drags down our learning efficiency. When I was studying abroad, I often thought: if only there were a specialized vocabulary book for each subject!
Well, here it is! The most complete English-Chinese vocabulary list for calculus is finally here! You're really catching it at the right time!
This vocabulary list is based on content from the following calculus topics:
Limits and Continuity: Definition and computation of limits; Infinite limits and asymptotes
Integrals: Definitions of definite and indefinite integrals; Basic integration techniques: substitution, integration by parts
...
Infinite Series :
Convergence tests for series; Power series and Taylor series
Multivariable Calculus: Partial derivatives and directional derivatives; Multiple integrals and vector calculus
To all students taking MATH10121, MATH10222, MATH20122, and MATH20411 at UoM — don't let calculus vocabulary slow down your progress. Save this vocabulary list now, learn efficiently, and enjoy solving problems! Like + follow, and if there's anything else you'd like to see, just DM me #UniversityofManchester #UoMstudyabroad #UKstudents #calculusQ&A #UoMstudents
Gold Score: 0

R1 Output:
<think> First, I need to evaluate the relevance between the user's search term and the document to be assessed, and assign a score based on the evaluation criteria. The search term is "What does infinite series mean?", and the user's intent is to understand the concept and meaning of infinite series.
Next, let's look at the content of the document. The title mentions "The Most Comprehensive Calculus Vocabulary List" and covers multiple topics from calculus courses, including limits and continuity, derivatives, integrals, infinite series, and multivariable calculus.
In the section on "Infinite Series," the document lists terms such as convergence tests and Taylor series. When analyzing relevance, the core of the search term is the definition and meaning of "infinite series." Although the document does mention content related to infinite series, it primarily presents a vocabulary list of related terms and methods, such as convergence tests and power series, without providing a detailed explanation or mathematical definition of the concept. The purpose of the document is to offer a vocabulary list to help students memorize and learn technical terms in calculus, rather than to explain each concept in depth.
According to the scoring criteria, a score of 2 requires that the document directly meets the main need of the search term and provides a complete answer. Since the user is looking for an explanation of infinite series and the document does not provide detailed content in this regard—only related terms and methods—it partially addresses the topic. **Therefore, it should be scored as 1 point.** </think>
<answer> 1 </answer>

R³A Output in the 2nd Round (Ours):
<think> First, I need to assess the relevance between the user's search query and the document to be evaluated, and assign a score based on the evaluation criteria. The search term is "What does infinite series mean?", and the user's need is to understand the definition or basic concept of infinite series.
Next, I examine the content of the document. Most of the document consists of a usage guide for a calculus vocabulary list and the source courses. In the "Infinite Series" section, it mentions "tests for convergence of series" and "power series and Taylor series." These may relate to certain evaluation methods and topics within infinite series, **but they do not provide a definition or basic conceptual explanation of infinite series.**
According to the scoring criteria, a score of 2 requires that the document directly satisfies the main need of the search term and provides a complete answer. The search term calls for a definition or basic concept of infinite series, which the document does not provide. A score of 1 is appropriate when the document partially answers the query or under specific conditions. In this case, **the document does not provide a definition or basic concept of infinite series** , but instead refers to some related methods and topics such as convergence tests and power series. Therefore, the document does not directly answer the search term and does not provide a complete response.
As a result, the document is irrelevant to the main intent of the search term and should be given a score of 0. </think>
<extract> None </extract>
<score> 0 </score>

Table 5: Example 1: The case comparison using the R1 and R³A models. The highlighted content indicate the portions in the original text that overlap with the query keywords, as well as the key reasoning components generated by the model.

Example 2
Query: **Why haven't the tuition fees for Suzhou kindergartens been paid yet?**
Document: **Suzhou kindergartens are all free!!! It seems Suzhou is the only city in Jiangsu province with free kindergartens across the entire province.** The policy is great, saving us money! Every time we cross the street, buses stop far away to let us pass, so I have to drag my son along, afraid we'll be a step too slow.
Jiangsu Province's free kindergarten policy will be gradually implemented starting from the fall semester of 2025.
The specific implementation time and scope are as follows:
2025: Kindergarten for the youngest children will be free.
2026: Kindergarten for the youngest and middle-aged children will be free.
2027: Kindergarten for the youngest, middle-aged, and oldest children will be completely free.
The implementation areas include:
Nanjing City: Pukou District, Qixia District, Yuhuatai District.
Wuxi City: Xishan District, Huishan District, Jiangyin City.
Xuzhou City: Xuzhou Economic and Technological Development Zone.
Changzhou City: Xinbei District, Liyang City.
Suzhou City: Gusu District, Wuzhong District, Xiangcheng District, Changshu City, Taicang City, Suzhou Industrial Economic and Technological Development Zone.
...
Gold Score: 2

R³A Output (two rounds)
<intent> The user is inquiring about the reason for the absence of tuition fee collection in Suzhou. This likely involves policy changes or implementation delays. The intent is to confirm if the policy is now "free of charge" and understand the specific implementation details. </intent>
<think> The user wants to know why fees haven't been collected. The document directly addresses this by stating that kindergartens in Suzhou are now free and provides a detailed implementation timeline for different grades (2025-2027) and specific districts. Since the document directly satisfies the information need without requiring additional inference, it is highly relevant. </think>
<extract> **Suzhou kindergartens are all free!!! It seems Suzhou is the only city in Jiangsu with full-region fee waivers.** </extract>
<score> 2 </score>

Example 3
Query: **How to buy Huangshan tickets for those over 60?**
Document: Must-read for coming to Huangshan, get the free and half-price activities in one post!!
January Huangshan Scenic Area Ticket Price:
Winter Tour Ticket Price: Winter Tour: Full ticket 150 yuan each, half ticket 75 yuan each.
Winter Tour Ropeway Ticket Price: Winter Tour: Yungu and Taiping Ropeway tickets, 65 yuan each way; Yuping Ropeway ticket, 75 yuan each way; Xihai Grand Canyon Sightseeing Cable Car ticket, 80 yuan each way.
January Huangshan Scenic Area 4 Major Privilege Activities:
Activity: Free Ticket Opening Day Activity.
Activity time: Every Wednesday in the first quarter of 2024 (non-holidays), 12 days in total.
Reservation process: 1. Tourists enter the "Huangshan Tourism Official Platform" Alipay or WeChat mini-program, click the "Free Ticket Opening Day" activity carousel on the homepage
...
Activity: Free admission for youths under 18 and people aged 60 and above.
Activity content: **In the first quarter of 2024, youths under 18 and people aged 60 and above can enjoy free admission to Huangshan Scenic Area. Enter the Huangshan Tourism Official Platform mini-program Huangshan Scenic Area ticket reservation page, select the ticket type "Huangshan Ticket Free Consumption Voucher for Groups over 60" for reservation.**
Gold Score: 1

R³A Output (two rounds)
<intent> Purchasing methods or steps for Huangshan scenic area tickets specifically for people over 60 years old. </intent>
<think> The user's need is the purchasing methods or steps for Huangshan tickets for people over 60 years old.The document provides the reservation process, including how to reserve through the Huangshan Tourism Official Platform mini-program and select the corresponding ticket type for reservation. Although the document primarily focuses on the free ticket activity, it has provided steps and methods for obtaining tickets, even if it is a free ticket case. However, the search term may also include purchasing tickets in non-free ticket situations (e.g., standard price purchase), but there is no other relevant information in the document. Therefore, **the document partially answers the question of how to buy tickets but may not cover all possible situations or detailed steps outside the promotional period, so the score should be 1**. </think>
<extract> In the first quarter of 2024, youths under 18 and people aged 60 and above can enjoy free admission to Huangshan Scenic Area. Enter the Huangshan Tourism Official Platform mini-program Huangshan Scenic Area ticket reservation page, select the ticket type "Huangshan Ticket Free Consumption Voucher for Groups over 60" for reservation. </extract>
<score> 1 </score>

Table 6: Example 2&3: Case study of R³A's performance on test set of NoteRel dataset.