

000 001 002 003 004 005 006 007 008 009 010 THE OVERTHINKING PREDICAMENT: WHEN REASON- ING HURTS RANKING

005 **Anonymous authors**

006 Paper under double-blind review

009 ABSTRACT

011 Document reranking is a key component in information retrieval (IR), aimed at
 012 refining initial retrieval results to improve ranking quality for downstream tasks.
 013 Recent studies—motivated by large reasoning models (LRMs)—have begun in-
 014 corporating explicit chain-of-thought (CoT) reasoning into LLM-based rerankers.
 015 However, the effectiveness of such reasoning for ranking tasks remains underex-
 016 plored. In this work, we present the first systematic study of reasoning in reranking
 017 across both **logits-based pointwise** and listwise settings, under both supervised
 018 fine-tuning and reinforcement learning. Using diverse benchmarks, including
 019 reasoning-intensive datasets (BRIGHT) and standard IR benchmarks (BEIR), we
 020 find that *reasoning-augmented rerankers consistently underperform their direct*
 021 *counterparts that predict rankings without CoT*, despite substantially higher infer-
 022 ence costs. Our analysis reveals three core limitations: (i) in pointwise rerankers,
 023 reasoning breaks calibration and biases models toward the positive class, rais-
 024 ing TPR but lowering TNR, which inflates false positives and degrades ranking in
 025 negative-dominant pools; (ii) **in listwise rerankers, explicit reasoning improves the**
 026 **fit during training but leads to higher variance and fails to improve performance in**
 027 **both in-domain and out-of-domain evaluations, even when reinforcement learning**
 028 **shortens rationales**; and (iii) overall, directly fine-tuned rerankers remain more
 029 stable, effective, and robust. These findings challenge the assumption that ex-
 030 plicit reasoning is universally beneficial for reranking. We conclude by highlight-
 031 ing future directions, including calibration-aware scoring for pointwise rerankers
 032 and the design of concise, targeted reasoning strategies to mitigate overfitting and
 033 overthinking in listwise rerankers.

034 1 INTRODUCTION

035 Document reranking is a crucial step in information retrieval (IR), aimed at refining the coarse-
 036 grained results produced by first-stage retrieval methods. By reordering candidate documents,
 037 reranking improves precision and overall ranking quality, which is essential for downstream applica-
 038 tions such as retrieval-augmented generation (RAG) (Lewis et al., 2020) and recommendation (Ren
 039 et al., 2024). The landscape of reranking is dominated by two primary paradigms: *pointwise* and *list-
 040 wise*. Pointwise rerankers independently estimate the relevance score of each query–document pair
 041 and sort documents accordingly. Since each document is processed in isolation, pointwise rerankers
 042 allow parallel computation and efficiency. In contrast, listwise rerankers consider the entire candi-
 043 date set jointly, asking the model to output a ranked list. While computationally more expensive,
 044 listwise rerankers often achieve more accurate rankings by leveraging cross-document interactions
 045 and relative comparisons, which is fundamentally easier than assigning a precise relevance score to
 046 each document in isolation.

047 With the rise of large language models (LLMs), reranking performance has advanced substantially.
 048 By combining targeted prompts with task-specific fine-tuning, LLM-based rerankers have achieved
 049 state-of-the-art results on diverse benchmarks (Sun et al., 2023a). Recently, large reasoning models
 050 (LRMs), such as DeepSeek-R1 (Guo et al., 2025) and OpenAI o1 (Jaech et al., 2024), have further
 051 drawn attention. Unlike typical LLMs that directly produce answers, LRMs explicitly decode rea-
 052 soning chains before providing the final prediction. This process narrows the gap between input and
 053 output, smooths token-by-token transitions, and has been shown to improve performance in many
 tasks. Motivated by these advances, recent studies have sought to extend test-time reasoning to

054 reranking via supervised fine-tuning Weller et al. (2025); Ji et al. (2025); Yang et al. (2025) or using
 055 reinforcement learning Zhang et al. (2025); Zhuang et al. (2025); Liu et al. (2025).
 056

057 Despite these developments, a fundamental question remains unresolved: **Does explicit reasoning**
 058 **truly benefit reranking?** Prior work often assumes that chain-of-thought (CoT) reasoning enhances
 059 reranking (Weller et al., 2025; Yang et al., 2025; Zhuang et al., 2025), yet such claims are rarely
 060 supported by fair comparisons against non-reasoning baselines (Weller et al., 2025; Yang et al.,
 061 2025; Zhuang et al., 2025; Liu et al., 2025). Moreover, emerging evidence suggests that reasoning-
 062 augmented rerankers can suffer from *overthinking* and lengthy reasoning chains introduce noise that
 063 degrades performance (Jedidi et al., 2025; Fan et al., 2025). However, these analyses are limited in
 064 scope, focusing narrowly on pointwise rerankers trained with supervised objectives, and fail to offer
 065 a systematic understanding of reasoning’s actual role in reranking.

066 In this work, we present the *first comprehensive and fair study of reasoning in reranking*. To ensure
 067 rigor and comparability, we adopt a unified experimental design: all rerankers are trained on the
 068 MS MARCO dataset, with reasoning-augmented models using CoT chains generated by DeepSeek-
 069 R1. We cover both **logits-based pointwise** and **listwise** rerankers, both *direct-output* and *reasoning-
 070 augmented* variants, and both *supervised fine-tuning (SFT)* and *reinforcement learning (RL)* training
 071 regimes. This setup eliminates inconsistencies across prior work and allows for a clean, apples-
 072 to-apples comparison. We further evaluate models on two complementary benchmarks: BRIGHT,
 073 which emphasizes reasoning-intensive queries, and BEIR, a standard suite of retrieval datasets. The
 074 scale, diversity, and uniformity of this design ensure that our conclusions are not anecdotal but
 075 systematically validated. Our extensive experiments reveal a striking and consistent pattern: **under-
 076 current training and inference setups, reasoning-based rerankers underperform their direct-output**
 077 **counterparts, even though they incur substantially higher inference costs.** This finding holds across
 078 architectures, training strategies, and benchmarks, suggesting that explicit reasoning—which ben-
 079 efits many other LLM tasks—does not translate into gains for reranking. Instead, reasoning intro-
 080 duces calibration errors, overthinking, and poor generalization, ultimately harming ranking quality.
 081 Our contributions can be summarized as follows:
 082

- 083 • **A rigorous, systematic study.** We conduct the first large-scale, controlled comparison of rea-
 084 soning vs. direct reranking, covering pointwise and listwise paradigms, SFT and RL training,
 085 and both reasoning-intensive and standard IR benchmarks.
- 086 • **Clear evidence against reasoning in reranking.** Direct-output rerankers consistently outper-
 087 form reasoning-augmented variants, despite the latter’s substantially higher inference cost.
- 088 • **Deeper insights into failure modes.** Our analysis reveals that for pointwise rerankers, rea-
 089 soning does not improve calibrated relevance prediction; instead, it shifts the error distribu-
 090 tion—raising TPR while reducing TNR—which disrupts score calibration and introduces a bias
 091 toward false positives. Similarly, for listwise rerankers, **reasoning leads to better training fit but**
 092 **increases variance and fails to yield gains on both in-domain and out-of-domain evaluations**,
 093 even when rationales are shortened via GRPO.
- 094 • **Guidance for future research.** Our findings suggest that reranking should prioritize effi-
 095 cient direct scoring rather than reasoning-heavy approaches. Promising directions include
 096 calibration-aware scoring for pointwise rerankers and designing concise, targeted reasoning
 097 strategies to mitigate overfitting and overthinking in listwise rerankers.

098 2 PRELIMINARIES

099 2.1 TASK SETUP

100 We consider the *reranking task* in information retrieval (IR), where the goal is to reorder an initial
 101 set of candidate documents so that those most relevant to a query appear at the top. Formally, given
 102 a query q , a retriever first returns a candidate set
 103

$$104 C(q) = \{d_1, d_2, \dots, d_k\}.$$

105 The reranker then takes $(q, C(q))$ as input and produces an improved ordering of the documents in
 106 $C(q)$. This approach reflects the standard two-stage retrieve-and-rerank architecture used in modern
 107 IR systems. First, an efficient retriever, optimized for recall, generates an initial, coarsely-ranked list

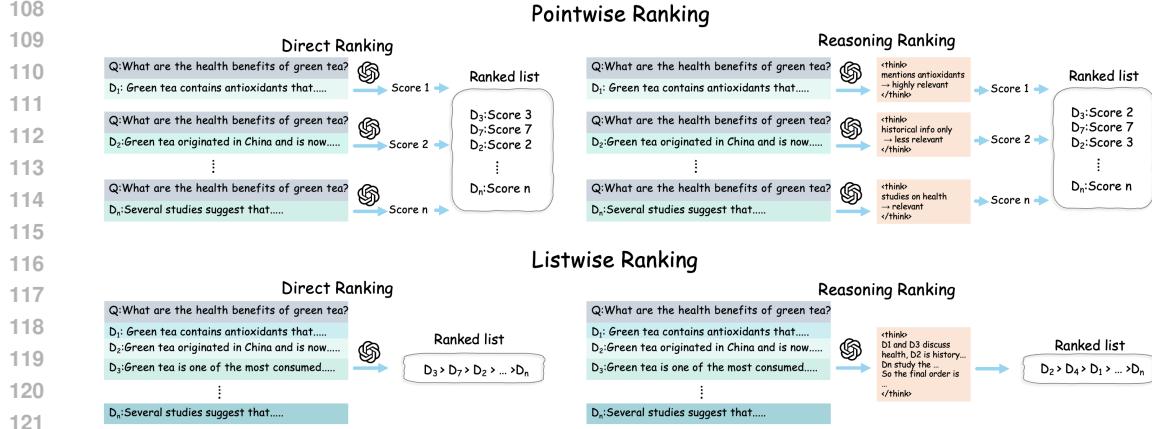


Figure 1: Illustration of Pointwise and Listwise Reranking (Direct vs. Reasoning). In pointwise, each query–document pair is judged independently, with relevance scores computed as the normalized probability of the TRUE token over {TRUE, FALSE} logits. Listwise directly optimizes the ranking order over candidate sets, with or without explicit reasoning.

of documents from a large corpus. Then, a reranker, typically a more expressive model optimized for precision, refines this list. The goal of this two-stage process is to maximize user satisfaction by placing the most relevant documents at the very top of the results.

2.2 POINTWISE RERANKER

In the pointwise setting, each query–document pair (q, d_i) is evaluated independently. Let $\xi(q, d_i)$ denote the prompt that encodes the pair, and let the answer space be $\mathcal{A} = \{\text{TRUE}, \text{FALSE}\}$, corresponding to tokens τ_{TRUE} and τ_{FALSE} . For candidate d_i , the model produces logits $\ell_i \in \mathbb{R}^{|\mathcal{V}|}$, from which a relevance score is derived as the normalized probability of the TRUE token:

$$s_i = \frac{\exp(\ell_i[\tau_{\text{TRUE}}])}{\exp(\ell_i[\tau_{\text{TRUE}}]) + \exp(\ell_i[\tau_{\text{FALSE}}])}.$$

The final ranking is obtained by sorting $\{s_i\}_{i=1}^k$ in descending order.

Optionally, a pointwise reranker can be extended with explicit reasoning: before predicting the binary decision, the model generates an intermediate reasoning trace z_i (e.g., a chain-of-thought),

$$z_i \sim p_\theta(z \mid \xi(q, d_i)), \quad s_i = \Pr_\theta(a = \text{TRUE} \mid \xi(q, d_i), z_i),$$

where $a \in \mathcal{A}$ and the probability is computed from the answer-token distribution conditioned on z_i . In practice, multiple traces $\{z_i^{(m)}\}_{m=1}^M$ may be sampled and aggregated (e.g., by averaging or voting). Suppressing z_i recovers the non-reasoning formulation above.

2.3 LISTWISE RERANKER

In the listwise setting, the model considers the entire candidate set $C(q)$ jointly. Let $\varphi(q, C(q))$ denote the encoding of the query and its candidate list. The model autoregressively generates a permutation $\pi = \langle \pi_1, \dots, \pi_k \rangle$ of indices:

$$\pi \sim p_\theta(\pi \mid \varphi(q, C(q))), \quad \pi_j \in \{1, \dots, k\} \setminus \{\pi_1, \dots, \pi_{j-1}\}.$$

At inference, one may decode the most likely permutation $\hat{\pi}$ (e.g., via greedy or beam search), or compute ranking scores from partial sequence probabilities. When k exceeds the context window, candidates can be processed in overlapping blocks (e.g., sliding windows), with local rankings merged into a global order.

Optionally, a listwise reranker can be extended with explicit reasoning. In this case, the model first generates a global reasoning trace Z that captures cross-document comparisons:

$$Z \sim p_\theta(Z \mid \varphi(q, C(q))), \quad \pi \sim p_\theta(\pi \mid \varphi(q, C(q)), Z).$$

162 The trace Z may include pairwise judgments, list-level critiques, or structured deliberation, and can
 163 be produced either as a separate stage (generate Z then π) or interleaved with ranking. Suppressing
 164 Z recovers the standard listwise formulation described above.

166 3 EVALUATING THE IMPACT OF REASONING ON RERANKING

168 3.1 MODEL VARIANTS

170 We design four LLM-based rerankers, each corresponding to a different combination of *pointwise*
 171 *vs. listwise* and *reasoning vs. non-reasoning* paradigms:

- 173 • **Direct-Point** (Non-Reasoning Pointwise): the model directly outputs a binary relevance
 174 decision (TRUE/FALSE). We take the logits of the answer token and transform them into a
 175 probability score, which is used for ranking.
- 176 • **Reason-Point** (Reasoning Pointwise): the model first generates a reasoning trace describ-
 177 ing why the document may or may not be relevant, and then produces the final binary
 178 decision. The relevance score is computed from the logits at the answer token position.
- 179 • **Direct-List** (Non-Reasoning Listwise): the model takes the entire candidate list as input
 180 and directly generates a permutation as the output ranking, e.g., $[3] > [5] > [4] > \dots$.
- 181 • **Reason-List** (Reasoning Listwise): the model first generates a reasoning sequence that
 182 compares and analyzes candidates, and then outputs the final ranking sequence.

184 3.2 TRAINING DETAILS

186 **Backbone Models** We adopt the Qwen3 series as the backbone for our rerankers, specifically
 187 Qwen3-4B and Qwen3-8B. This choice is consistent with the prevailing practice in the reranking
 188 community¹. All training experiments are conducted on two NVIDIA A100 (80 GB) GPUs.

189 **Pointwise Rerankers** We study two pointwise variants: *Direct-Point* and *Reason-Point*. Both
 190 models are trained on the RANK1 corpus (Weller et al., 2025) derived from MS MARCO, comprising
 191 $\sim 386k$ query–passage pairs annotated by DeepSeek-R1 with a chain-of-thought rationale and
 192 a binary answer (TRUE/FALSE). For *Reason-Point*, we perform supervised fine-tuning on quadru-
 193 ples \langle query, passage, rationale, answer \rangle . For *Direct-Point*, we ablate the rationale and fine-tuning on
 194 \langle query, passage, answer \rangle , training the model to emit a single token in {TRUE, FALSE}. Both variants
 195 minimize cross-entropy loss. We employ the LLaMA-Factory² framework for supervised fine-
 196 tuning. All models are trained using LoRA with rank 32 and $\alpha = 64$, a learning rate of 1×10^{-4} , and
 197 cross-entropy loss. Both DIRECT-POINT and REASON-POINT rerankers are trained for one epoch.
 198 At inference time, we compute the relevance score used for ranking as the probability assigned to
 199 TRUE via a two-way softmax over the logits of {TRUE, FALSE}; ties are broken by the logit margin.
 200 Example prompts and data instances for both settings are provided in the Appendix C.1.

201 **Listwise Rerankers** We train listwise rerankers on the REASONRANK training corpus (Liu et al.,
 202 2025), which contains $\sim 13k$ query–candidate sets primarily derived from MS MARCO and related
 203 benchmarks, split evenly between SFT and GRPO (approximately 6.7k each). Each instance in-
 204 cludes a query, a candidate set, a rationale produced with DeepSeek-R1, and a gold ranking order.
 205 We consider two variants: *Direct-List*, which generates an ordering directly from the query and can-
 206 didate set, and *Reason-List*, which is prompted to first generate own rationale and then produce the
 207 final ordering.

208 We adopt two-stage training for training *Direct-List* and *Reason-List*. Stage 1 performs super-
 209 vised fine-tuning (SFT) to teach the model to output ranking sequences. To encourage struc-
 210 turally valid outputs, prompts require the model to produce a reasoning segment demarcated by
 211 \langle think $\rangle \dots \langle$ think \rangle and a final ranking in \langle answer $\rangle \dots \langle$ answer \rangle format.
 212 Stage 2 refines the SFT model with Group Relative Policy Optimization (GRPO) (Guo et al., 2025).

213
 214 ¹A complete list of backbone models used in both our work and prior reasoning-enhanced rerankers is
 215 provided in Appendix A.

216 ²<https://github.com/hiyouga/LLaMA-Factory>

We follow the setting in ReasonRank (Liu et al., 2025), using a composite *multi-view* ranking reward that reflects position sensitivity, coverage, and list similarity:

Let y^{list} denote the predicted ranking sequence and y' the gold ranking. We combine three signals:

$$R_m = \text{NDCG}@10(y^{\text{list}}, y') + \phi \cdot \text{Recall}@10(y^{\text{list}}, y') + \gamma \cdot \text{RBO}(y^{\text{list}}, y'), \quad (1)$$

where ϕ, γ weight coverage and overlap. Rank-Biased Overlap (RBO) (Webber et al., 2010) emphasizes top ranks and is computed as

$$\text{RBO}(y^{\text{list}}, y') = (1 - p) \sum_{d=1}^{|y^{\text{list}}|} p^{d-1} \frac{|y_{1:d}^{\text{list}} \cap y'_{1:d}|}{d}, \quad (2)$$

with persistence parameter $p \in (0, 1)$ and $y_{1:d}$ the top- d prefix. Following REASONRANK, we gate R_m with simple format validators to stabilize learning:

$$R = \begin{cases} R_m, & \text{both output and answer formats are valid,} \\ 0, & \text{only the output format is valid,} \\ -1, & \text{otherwise,} \end{cases} \quad (3)$$

where the *output-format* check requires the presence of `<think>` and `<answer>` tags, and the *answer-format* check verifies that the `<answer>` contains a canonical listwise ordering.

With the gated multi-view reward in Eqs. (1)–(2), we refine the SFT policy via GRPO (Guo et al., 2025). Given input x , a group of samples $\mathcal{G} = \{y_i\}$ is drawn, sequence rewards $R(x, y_i)$ are converted to token-level advantages $\hat{A}_{i,t}$, and the policy is updated by the clipped objective:

$$\mathcal{J}_{\text{GRPO}}(\theta) = -\frac{1}{|\mathcal{G}|} \sum_{i,t} \min(r_{i,t}(\theta) \hat{A}_{i,t}, \text{clip}(r_{i,t}(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_{i,t}) - \beta D_{\text{KL}}(\pi_\theta \parallel \pi_{\text{ref}}), \quad (4)$$

where $r_{i,t}(\theta)$ is the importance ratio and π_{ref} the SFT reference. We implement GRPO with Ver1³ for one epoch on the ReasonRank data, producing *Direct-List* and *Reason-List* models.

3.3 EXPERIMENTAL SETUP

All experiments are conducted on two NVIDIA A100 (80 GB) GPUs. Rerankers operate on a fixed first-stage candidate pool of $k=100$ passages per query built with BM25. **Unified system prompts are used across all datasets, as shown in Appendix C. Dataset-level instructions strictly follow the original Rank1 templates** (Weller et al., 2025).

Baselines We compare our four proposed rerankers—*Direct-Point*, *Reason-Point*, *Direct-List*, and *Reason-List*—including ablations across training stages (SFT only vs. SFT+GRPO). In addition, we compare our models against state-of-the-art *reasoning-enhanced* LLM rerankers from prior work: Pointwise: Rank1-7B, Rank1-14B (Weller et al., 2025), TF-Rank-4B, TF-Rank-8B (Fan et al., 2025); Listwise: Rank-R1-7B, Rank-R1-14B (Zhuang et al., 2025), REARank-7B (Zhang et al., 2025), and ReasonRank-7B (Liu et al., 2025).

Benchmarks and Metrics We evaluate on two retrieval benchmarks: BRIGHT, a reasoning-intensive IR suite spanning diverse domains, and BEIR, a standard heterogeneous IR benchmark. Following common practice, the primary metric is NDCG@10, which captures both relevance and position sensitivity on the top-10 results.

3.4 MAIN RESULTS

Reasoning Does Not Improve Reranking Performance As shown in Tables 1 and 2, a consistent and repeatable pattern emerges across *all* training settings (SFT and SFT+GRPO), model sizes (4B/8B), and benchmarks (BRIGHT and BEIR): **direct rerankers consistently outperform their reasoning-augmented counterparts**. For pointwise rerankers on BRIGHT, *Direct-Point-4B*

³<https://github.com/volcengine/ver1>

270 exceeds *Reason-Point-4B* by $\Delta N@10 = +9.0$, and *Direct-Point-8B* by $+6.1$ on the original query,
 271 while the advantage remains $+4\text{--}6$ points on the gpt4_reason query split.⁴; on BEIR, the corresponding
 272 gaps are $+5.3$ and $+4.3$. For listwise rerankers, the advantage is smaller but remains stable: on
 273 BRIGHT, *Direct-List* achieves $+0.4\text{--}+1.7$ higher $N@10$ than *Reason-List*, while on BEIR the margin
 274 ranges from $+0.3$ to $+1.9$, consistently observed under both SFT and SFT+GRPO. Moreover,
 275 GRPO provides further improvements for listwise rerankers compared to SFT alone, yet the su-
 276 periority of direct reranking persists regardless of training strategy. These results reveal one clear
 277 trend: under current logits-based pointwise and generative listwise reranking setups, we observe that
 278 explicit reasoning does *not* improve ranking performance.

279 **Comparison Results on BRIGHT and BEIR.** Tables 1 and 2 report the performance of our
 280 rerankers compared to reasoning-enhanced baselines on BRIGHT and BEIR, respectively. On orig-
 281 inal query split of BRIGHT, our *Direct-List-8B* achieves the best overall score with $N@10 = 27.1$,
 282 followed closely by *Direct-Point-8B* at 26.8, both outperforming reasoning-based baselines such as
 283 *ReasonRank-7B* (26.4), *TFRank-8B* (22.6), and *Rank-R1-14B* (20.5). On the gpt4_reason split of
 284 BRIGHT, *Direct-List-4B* and *Direct-List-8B* (both $N@10 = 35.3$) likewise surpass all Reason-List
 285 baselines. On BEIR, the strongest model is *Direct-Point-4B*, which obtains $N@10 = 45.4$, surpassing
 286 larger reasoning-enhanced listwise rerankers such as *Rank-R1-14B* (43.8) and *ReasonRank-7B*
 287 (41.7). These results demonstrate that our *direct* rerankers not only outperform existing reasoning-
 288 based counterparts, but also highlight that explicit reasoning is unnecessary for achieving state-of-
 289 the-art effectiveness in LLM-based reranking.

290 Table 1: Performance comparison on BRIGHT across different reranker variants. We report results
 291 for Direct-Point, Reason-Point, Direct-List, and Reason-List under both SFT and GRPO training,
 292 together with representative pointwise and listwise baselines.

Model	Training	StackExchange						Coding			Theorem-based			Avg.
		Bio.	Earth.	Econ.	Psy.	Rob.	Stack.	Sus.	Leet.	Pony	AoPS	TheoQ.	TheoT.	
Pointwise														
BM25	/	18.9	27.2	14.9	12.5	13.6	18.4	15.0	7.9	24.4	6.2	4.9	10.4	14.5
Rank1-7B	SFT	31.4	36.7	18.3	25.4	13.8	17.6	24.8	16.7	9.5	6.1	9.5	11.6	18.5
Rank1-14B	SFT	29.6	34.8	17.2	24.3	18.6	16.2	24.5	17.5	14.4	5.5	9.2	10.7	18.5
TFRank-4B	SFT+GRPO	33.2	45.9	17.6	29.5	21.0	20.9	18.3	25.0	9.1	9.5	9.8	7.3	20.6
TFRank-8B	SFT+GRPO	33.7	46.2	23.7	26.0	24.1	20.1	23.6	28.8	12.5	10.8	11.4	9.7	22.6
Reason-Point-4B	SFT	23.6	29.0	15.0	23.7	16.7	12.2	18.3	18.4	12.4	8.9	11.0	9.4	16.5
Direct-Point-4B	SFT	34.9	45.1	23.3	31.8	26.6	23.6	30.7	18.5	35.4	7.2	13.6	15.2	25.5
Reason-Point-8B	SFT	24.9	34.6	17.5	26.2	25.9	22.4	19.7	11.9	36.6	9.3	6.5	12.6	20.7
Direct-Point-8B	SFT	33.9	46.4	24.6	31.6	25.8	25.9	32.0	25.3	35.5	12.0	13.5	15.2	26.8
Listwise														
Rank-R1-7B	GRPO	26.0	28.5	17.2	24.2	19.1	10.4	24.2	19.8	4.3	4.3	8.3	10.9	16.4
Rank-R1-14B	GRPO	31.2	38.5	21.2	26.4	22.6	18.9	27.5	20.2	9.2	9.7	9.2	11.9	20.5
REARANK-7B	GRPO	23.4	27.4	18.5	24.2	17.4	16.3	25.1	27.0	8.0	7.4	7.9	9.5	17.7
ReasonRank-7B	SFT+GRPO	36.3	44.2	24.8	31.7	30.7	24.9	32.8	28.7	17.5	12.0	18.5	14.0	26.4
Reason-List-4B	SFT	30.7	37.3	18.7	27.7	27.9	19.8	28.5	28.1	13.7	9.1	13.9	13.3	22.4
Direct-List-4B	SFT	32.7	38.6	20.0	28.4	28.6	20.5	31.2	30.9	15.1	10.4	17.8	15.6	24.1
Reason-List-8B	SFT	31.9	39.6	22.4	29.0	29.9	23.4	34.5	26.8	18.9	9.7	15.6	12.1	24.5
Direct-List-8B	SFT	32.6	38.4	21.3	28.9	31.9	22.6	31.8	28.9	16.9	11.1	18.5	15.4	24.9
Reason-List-4B	SFT+GRPO	33.6	40.8	21.6	28.0	33.3	26.0	29.3	31.0	13.3	11.4	16.5	15.4	25.0
Direct-List-4B	SFT+GRPO	33.8	41.5	23.4	29.3	34.0	23.9	34.2	33.4	13.7	11.9	17.1	14.6	25.9
Reason-List-8B	SFT+GRPO	32.1	40.3	26.7	32.1	30.0	25.5	33.8	28.8	19.4	9.8	18.0	14.0	25.9
Direct-List-8B	SFT+GRPO	35.2	42.7	23.1	30.6	34.0	27.6	33.9	29.2	22.9	12.1	17.9	15.8	27.1

4 ANALYSIS

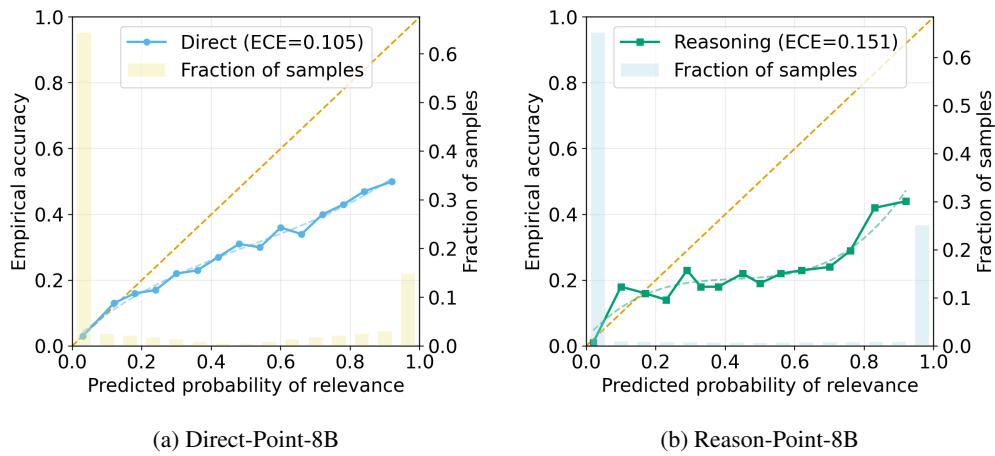
4.1 POINTWISE: CALIBRATION FAILURE AND TRUE BIAS WITH REASONING

321 **Reasoning breaks calibration of confidence and accuracy** Calibration assesses whether pre-
 322 dicted probabilities match the true likelihood of relevance. In pointwise rerankers, the score as-

323 ⁴Complete results for the gpt4_reason split are reported in Appendix B.

324
 325 Table 2: Performance comparison on BEIR across different reranker variants. We report results
 326 for Direct-Point, Reason-Point, Direct-List, and Reason-List under both SFT and GRPO training,
 327 together with representative pointwise and listwise baselines

328 Model	329 Training	ArguA	ClimF	DBP	FiQA	NFCorp	SciDoc	SciFact	Touche	TrecC	Avg.
Pointwise											
Listwise											
BM25	/	39.7	16.5	31.8	23.6	33.8	14.9	67.9	44.2	59.5	36.9
Rank1-7B	SFT	26.4	16.2	37.7	38.4	37.9	16.5	76.1	24.5	79.5	39.2
Rank1-14B	SFT	32.2	15.6	34.2	36.6	35.1	16.6	73.8	25.9	78.0	38.7
TF-Rank-4B	SFT+GRPO	37.2	19.7	37.9	36.2	38.3	18.3	76.6	37.6	80.5	42.5
TF-Rank-8B	SFT+GRPO	36.5	21.7	37.0	38.0	38.0	17.9	74.6	35.0	80.0	42.1
Reason-Point-4B	SFT	38.1	16.5	39.1	36.0	30.8	16.8	74.9	27.1	81.3	40.1
Direct-Point-4B	SFT	57.5	19.6	43.4	42.4	35.9	18.6	77.4	30.6	83.5	45.4
Reason-Point-8B	SFT	40.1	13.8	37.5	35.6	32.5	18.2	74.1	21.6	80.3	39.3
Direct-Point-8B	SFT	58.6	15.7	42.7	43.3	36.2	16.8	75.3	22.3	82.7	<u>43.7</u>
Listwise											
Rank-R1-7B	GRPO	37.0	24.1	43.2	40.1	36.2	18.8	76.1	33.0	82.6	<u>43.5</u>
Rank-R1-14B	GRPO	34.4	24.2	44.0	43.0	37.9	19.7	77.5	29.6	83.9	43.8
REARANK-7B	GRPO	35.6	20.6	43.5	35.8	37.9	19.2	71.9	40.2	80.1	42.8
ReasonRank-7B	SFT+GRPO	33.3	20.0	44.7	38.2	36.6	19.7	72.8	30.4	79.6	41.7
Reason-List-4B	SFT	39.3	13.8	37.7	32.9	30.2	16.0	69.1	24.6	79.5	38.1
Direct-List-4B	SFT	41.7	14.1	36.5	37.0	33.5	16.4	72.4	22.9	78.1	39.2
Reason-List-8B	SFT	32.8	16.2	42.2	36.6	36.0	18.1	69.7	27.1	79.2	39.8
Direct-List-8B	SFT	28.9	16.4	42.3	37.7	35.9	18.9	73.6	27.0	80.1	40.1
Reason-List-4B	SFT+GRPO	30.8	15.7	43.3	36.1	36.5	18.2	73.1	26.4	78.0	39.8
Direct-List-4B	SFT+GRPO	36.7	19.4	43.8	34.4	36.4	18.2	69.8	29.4	77.7	40.6
Reason-List-8B	SFT+GRPO	28.6	19.9	43.5	35.5	37.2	18.8	72.5	24.9	77.8	39.9
Direct-List-8B	SFT+GRPO	36.4	19.4	45.1	38.2	36.9	18.7	71.0	31.6	78.7	41.8



365 Figure 2: Calibration curves of pointwise rerankers: predicted probabilities vs. empirical accuracies.
 366

367 signed to a candidate is interpreted as the model’s confidence that it is relevant. A perfectly calibrated
 368 model satisfies, for example, that predictions of 0.9 correspond to roughly 90% truly relevant
 369 items; in reliability diagrams this appears as points along the diagonal $y=x$. To quantify deviations
 370 from perfect calibration, we use the *Expected Calibration Error* (ECE) (Guo et al., 2017),
 371 the weighted discrepancy between predicted confidence and empirical accuracy across M bins:
 372
$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{N} |\text{acc}(B_m) - \text{conf}(B_m)|$$
, where B_m is the set of samples in bin m , $|B_m|$
 373 its size, N the total number of samples, $\text{acc}(B_m)$ the empirical accuracy, and $\text{conf}(B_m)$ the average
 374 predicted probability in that bin. Smaller values indicate better calibration. Notably, we observe a
 375 pronounced polarization in the logits: most predictions cluster near 0 or 1, reflecting overconfident
 376 decision boundaries. As shown in Figure 2, based on results from the BEIR, the direct pointwise
 377

378
379
380
381 Table 3: Class-conditional performance on pointwise
382 rerankers. We report TPR (%) and TNR (%).
383

Model	Biology		MS MARCO		Avg.
	TPR	TNR	TPR	TNR	
DeepSeek-R1	52.4	96.1	40.8	85.1	68.6
Reason-Point-4B	43.7	91.3	38.7	79.4	63.3
Direct-Point-4B	34.0	93.2	30.7	85.7	60.9
Reason-Point-8B	50.5	98.1	35.9	85.5	67.5
Direct-Point-8B	31.1	100.0	25.5	94.2	62.7

384
385
386
387
388
389
390 reranker—though not perfect—maintains a clear monotonic relationship between confidence and
391 accuracy ($ECE = 0.105$). By contrast, the reasoning-enhanced reranker exhibits systematic over-
392 confidence with larger departures from the diagonal ($ECE = 0.151$), indicating that *adding reasoning breaks confidence calibration* in the pointwise setting. This miscalibration helps explain the
393 observed degradation in ranking quality (e.g., lower NDCG).
394

395
396 **Reasoning increases “True” proclivity** Beyond aggregate calibration, class-conditional analy-
397 sis reveals a consistent shift toward predicting the positive class. The training data have a positive:negative ratio of approximately 1:2. To match this prior, we construct evaluation pools with
398 100 positives and 200 negatives per query—both in-domain (MS MARCO DL19/DL20) and out-
399 of-domain (BRIGHT–Biology)—and also include the teacher model (DeepSeek-R1). Letting the
400 reranker judge relevance and decode the answer token, we report class-conditional performance in
401 Table 3 using standard notation: TPR (true positive rate; recall on positives) and TNR (true negative
402 rate; specificity = $1 - FPR$). Under this matched prior, *Reason* models tend to achieve higher macro
403 binary accuracy (the mean of TPR and TNR) than *Direct* models; however, the gains consistently
404 arise from *higher TPR* coupled with *lower TNR* (i.e., higher FPR). In reranking regimes where neg-
405 atives dominate, this combination is detrimental: elevated FPR promotes non-relevant documents
406 into the head of the ranked list and, together with the calibration failure above, prevents the binary
407 accuracy gains from translating into improved ranking metrics.
408

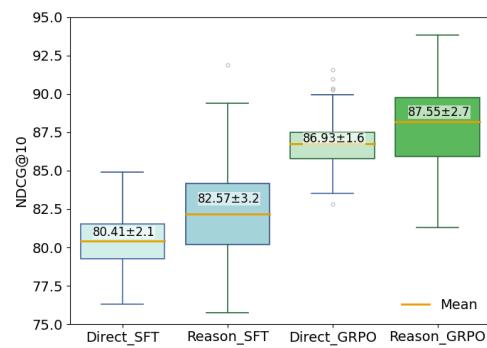
409
410

4.2 LISTWISE: REASONING IMPROVES TRAINING FIT BUT HURTS GENERALIZATION

411
412 **Reasoning boosts training fit but hurts gen-**
413 **eralization** Unlike pointwise rerankers that as-
414 sign scores to query–document pairs and then
415 sort, listwise objectives *directly optimize the per-*
416 *mutation* of a candidate set. We therefore ask
417 whether exposing chain-of-thought (CoT) helps
418 under listwise training. We evaluate four 8B
419 variants on a training split of 100 instances:
420 *Direct-List_SFT* vs. *Reason-List_SFT* and *Direct-*
421 *List_GRPO* vs. *Reason-List_GRPO*. As shown in
422 Fig. 3, reasoning attains higher mean NDCG@10
423 on the training split but with markedly larger dis-
424 persion: *Reason-List_SFT* 82.57 ± 3.2 vs. *Direct-*
425 *List_SFT* 80.41 ± 2.1 ($\Delta = +2.16$), and *Reason-*
426 *List_GRPO* 87.55 ± 2.7 vs. *Direct-List_GRPO*
427 86.93 ± 1.6 ($\Delta = +0.62$). These patterns indi-
428 ciate that CoT can better fit the target permutations
429 encountered during training, while simultaneously
430 introducing instance-level instability. On the in-
431 domain MS MARCO Dev sets (DL19/20), *Direct-List* consistently outperforms *Reason-List* across
432 both 4B and 8B backbones (Table 4). Concretely, *Direct-List-4B* surpasses its reasoning counterpart
433 by $+3.01$ (73.77 vs. 70.76) on DL19 and $+0.26$ (68.97 vs. 68.71) on DL20; *Direct-List-8B* leads by
434 $+0.40$ (73.00 vs. 72.60) on DL19 and $+1.57$ (71.38 vs. 69.81) on DL20. Thus, the training-split ad-
435 vantage of *Reason-List* does *not* translate to stronger in-domain performance. The same trend holds
436 on BRIGHT and BEIR: reasoning-based listwise models lag behind their direct counterparts, rein-

378
379
380
381 Table 4: Listwise (GRPO) performance
382 on MS MARCO (NDCG@10).
383

Model	MS MARCO	
	DL19	DL20
Direct-List-4B	73.77	68.97
Reason-List-4B	70.76	68.71
Direct-List-8B	73.00	71.38
Reason-List-8B	72.60	69.81

411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
Figure 3: Training-split listwise performance of
four 8B variants. Reasoning improves mean
NDCG@10 but increases variance.

437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
8010
8011
8012
8013
8014
8015
8016
8017
8018
8019
8020
8021
8022
8023
8024
8025
8026
8027
8028
8029
8030
8031
8032
8033
8034
8035
8036
8037
8038
8039
8040
8041
8042
8043
8044
8045
8046
8047
8048
8049
8050
8051
8052
8053
8054
8055
8056
8057
8058
8059
8060
8061
8062
8063
8064
8065
8066
8067
8068
8069
8070
8071
8072
8073
8074
8075
8076
8077
8078
8079
8080
8081
8082
8083
8084
8085
8086
8087
8088
8089
8090
8091
8092
8093
8094
8095
8096
8097
8098
8099
80100
80101
80102
80103
80104
80105
80106
80107
80108
80109
80110
80111
80112
80113
80114
80115
80116
80117
80118
80119
80120
80121
80122
80123
80124
80125
80126
80127
80128
80129
80130
80131
80132
80133
80134
80135
80136
80137
80138
80139
80140
80141
80142
80143
80144
80145
80146
80147
80148
80149
80150
80151
80152
80153
80154
80155
80156
80157
80158
80159
80160
80161
80162
80163
80164
80165
80166
80167
80168
80169
80170
80171
80172
80173
80174
80175
80176
80177
80178
80179
80180
80181
80182
80183
80184
80185
80186
80187
80188
80189
80190
80191
80192
80193
80194
80195
80196
80197
80198
80199
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262
80263
80264
80265
80266
80267
80268
80269
80270
80271
80272
80273
80274
80275
80276
80277
80278
80279
80280
80281
80282
80283
80284
80285
80286
80287
80288
80289
80290
80291
80292
80293
80294
80295
80296
80297
80298
80299
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262
80263
80264
80265
80266
80267
80268
80269
80270
80271
80272
80273
80274
80275
80276
80277
80278
80279
80280
80281
80282
80283
80284
80285
80286
80287
80288
80289
80290
80291
80292
80293
80294
80295
80296
80297
80298
80299
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262
80263
80264
80265
80266
80267
80268
80269
80270
80271
80272
80273
80274
80275
80276
80277
80278
80279
80280
80281
80282
80283
80284
80285
80286
80287
80288
80289
80290
80291
80292
80293
80294
80295
80296
80297
80298
80299
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262
80263
80264
80265
80266
80267
80268
80269
80270
80271
80272
80273
80274
80275
80276
80277
80278
80279
80280
80281
80282
80283
80284
80285
80286
80287
80288
80289
80290
80291
80292
80293
80294
80295
80296
80297
80298
80299
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262
80263
80264
80265
80266
80267
80268
80269
80270
80271
80272
80273
80274
80275
80276
80277
80278
80279
80280
80281
80282
80283
80284
80285
80286
80287
80288
80289
80290
80291
80292
80293
80294
80295
80296
80297
80298
80299
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262
80263
80264
80265
80266
80267
80268
80269
80270
80271
80272
80273
80274
80275
80276
80277
80278
80279
80280
80281
80282
80283
80284
80285
80286
80287
80288
80289
80290
80291
80292
80293
80294
80295
80296
80297
80298
80299
80200
80201
80202
80203
80204
80205
80206
80207
80208
80209
80210
80211
80212
80213
80214
80215
80216
80217
80218
80219
80220
80221
80222
80223
80224
80225
80226
80227
80228
80229
80230
80231
80232
80233
80234
80235
80236
80237
80238
80239
80240
80241
80242
80243
80244
80245
80246
80247
80248
80249
80250
80251
80252
80253
80254
80255
80256
80257
80258
80259
80260
80261
80262

432 forcing that CoT’s gains on the training split reflect improved *in-sample* fitting rather than genuine
 433 *out-of-domain* generalization.
 434

435 **GRPO improves performance and reduces overthinking.** As shown in Table 1, Table 2, and
 436 Fig. 3, GRPO yields substantial performance improvements over SFT training. At the same time,
 437 GRPO markedly shortens the rationales produced by reasoning models. On the training split (Fig. 3),
 438 the average rationale length decreases from 397.7 tokens in *Reason-List_SFT* to 172.3 in *Reason-List_GRPO*,
 439 reducing inference cost and mitigating extreme “overthinking,” yet achieving higher
 440 NDCG scores. This finding suggests that excessively long CoT rationales are not a prerequisite for
 441 producing effective ranking orders. The compression effect follows directly from the GRPO reward
 442 design (Eq. 3), which incentivizes only output format validity and ranking quality rather than ver-
 443 bose reasoning. Although GRPO enhances stability and efficiency, *Direct-List* models still achieve
 444 stronger generalization on MS MARCO DL19/20 and on BRIGHT/BEIR (Table 4), implying that
 445 shorter CoT mitigates overthinking but does not substitute for direct optimization of permutations.
 446 These observations point to a broader research direction: future work should explore how to design
 447 *concise and targeted* reasoning strategies that balance interpretability, stability, and generalization,
 448 while avoiding overfitting and reliance on lengthy CoT outputs.
 449

450 4.3 IMPLICATIONS FOR FUTURE RESEARCH

451 Our findings show that explicit reasoning does not inherently lead to performance gains in reranking,
 452 and therefore researchers should not assume that longer or more elaborate reasoning will universally
 453 improve ranking quality. Instead, the results point to two concrete directions for future work.

454 **For logits-based pointwise rerankers**, the performance degradation is primarily driven by score
 455 miscalibration rather than insufficient reasoning capacity. This suggests that future progress is more
 456 likely to come from *calibration-aware training objectives* that preserve score monotonicity, rather
 457 than increasing the depth or length of reasoning traces. Beyond calibration, another promising
 458 direction is to explore *token-based scoring and ranking mechanisms* (Shao et al., 2025; Fan et al.,
 459 2025), which operate at the token level rather than relying solely on a scalar logit. Such methods can
 460 produce more stable and fine-grained relevance signals, and therefore may complement calibration-
 461 oriented approaches in improving pointwise reranking.

462 **For listwise reranking**, we observe that the training data may contain unnecessarily long chain-of-
 463 thought traces, which increase inference cost without translating into better ranking quality. This
 464 suggests two promising directions: (i) developing *more concise or adaptive reasoning strategies* to
 465 reduce overthinking, and (ii) designing *reward formulations directly aligned with ranking metrics*,
 466 rather than relying on generic reasoning supervision.

467 Overall, direct models remain more stable and effective across both in-domain and OOD settings,
 468 highlighting *calibration* (for pointwise rerankers) and *objective alignment* (for listwise training) as
 469 the key bottlenecks for future research, rather than reasoning capacity itself.

470 5 RELATED WORKS

471 **LLMs for Ranking** The use of Large Language Models (LLMs) for ranking has emerged as a
 472 dominant paradigm in information retrieval, with methods that can be broadly classified into two
 473 primary approaches: *pointwise* and *listwise* (Qin et al., 2024; Lu et al., 2025). Pointwise rerankers
 474 evaluate the relevance of each query-document pair in isolation. This is typically achieved by fram-
 475 ing the task as a classification problem, where the model computes a relevance score from the output
 476 logits of binary tokens like “true” or “false”. This approach, exemplified by influential models such
 477 as MonoT5 (Nogueira et al., 2020), MonoBERT (Nogueira et al., 2019), and RankLLaMA (Ma
 478 et al., 2024), benefits from simplicity and computational efficiency, as each document can be scored
 479 independently. In contrast, the listwise paradigm is built on the principle of relative comparison,
 480 where multiple candidate documents are considered jointly to determine their final order. This cate-
 481 gory encompasses several implementation styles. The most direct application is generative listwise
 482 ranking, where models like RankGPT (Sun et al., 2023b), RankVicuna (Pradeep et al., 2023a), and
 483 RankZephyr (Pradeep et al., 2023b) leverage their generative capabilities to output a fully sorted
 484 list of documents. This broader paradigm also includes pairwise methods, which learn relative
 485

486 preferences by predicting the more relevant document from a pair (Qin et al., 2024), and setwise approaches that operate on a group of candidates, for example by identifying the single most relevant 487 document within the set (Zhuang et al., 2024). To manage the long input sequences inherent to this 488 approach, many listwise methods employ strategies like sliding windows or hierarchical ranking to 489 improve efficiency (Sharifymoghaddam et al., 2025).
490

492 **LRMs for Ranking** Inspired by the success of Large Reasoning Models (LRMs), a recent line 493 of research has focused on incorporating explicit reasoning into the reranking process to handle 494 complex queries and improve interpretability. This effort has largely followed two main technical 495 strategies: supervised fine-tuning with distillation and reinforcement learning. The first strategy 496 involves *Supervised Fine-Tuning (SFT) and distillation*, where reasoning capabilities are transferred 497 from powerful teacher models (e.g., GPT-4, DeepSeek-R1) to smaller, more efficient rerankers. For 498 instance, ReasoningRank (Ji et al., 2025) and Rank1 (Weller et al., 2025) distill pairwise or listwise 499 comparative rationales into models such as the LLaMA-3 and Qwen2.5 series. This approach has 500 also been extended to generative listwise settings, where reasoning tokens are directly integrated 501 into the ranking sequence (Yang et al., 2025). A complementary strategy employs *Reinforcement 502 Learning (RL)* to further refine these reasoning-aware models. Works like Rank-R1 (Zhuang et al., 503 2025) and REARANK (Zhang et al., 2025) use RL to directly optimize for ranking metrics, while 504 others such as TFRank (Fan et al., 2025) and ReasonRank (Liu et al., 2025) adopt a hybrid two- 505 stage approach combining SFT with subsequent RL fine-tuning. Beyond these dominant paradigms, 506 other methods have explored more structured formulations, such as modeling reranking as a decision 507 process to improve robustness (Lee et al., 2025; Niu et al., 2024). Despite these advances, the core 508 premise that explicit reasoning is beneficial is being called into question. Recent findings suggest 509 that for pointwise rerankers, the addition of reasoning can be detrimental, leading to issues like 510 overthinking (Jedidi et al., 2025; Fan et al., 2025). This emerging evidence highlights that the utility 511 of reasoning in reranking is far from settled, motivating the systematic investigation in our work.
512

513 6 CONCLUSION

514 In this work, we systematically examined the role of explicit reasoning in document reranking across 515 pointwise and listwise paradigms with SFT and GRPO training. Our findings are threefold: (i) 516 in pointwise rerankers, reasoning breaks calibration, yielding overconfident scores and degraded 517 ranking despite modest gains in binary accuracy; (ii) reasoning biases models toward the positive 518 class, raising TPR but reducing TNR, which is harmful in negative-dominant candidate pools; (iii) 519 in listwise rerankers, reasoning improves in-domain fit but increases variance and fails to generalize 520 out-of-domain, even when GRPO shortens rationales. Overall, direct models remain more stable 521 and effective, pointing to the need for calibration-aware objectives in pointwise rerankers and more 522 concise reasoning strategies in listwise rerankers to improve generalization.
523

524 ETHICS STATEMENT

525 This work does not involve human subjects, personal data, or sensitive information, and therefore 526 raises no direct ethical concerns. Our research focuses on reranking methods for information re- 527 trieval benchmarks, which do not include personally identifiable information or sensitive content. 528 We follow the ICLR Code of Ethics and ensure that our methodology and results are presented with 529 transparency and fairness.
530

533 534 REPRODUCIBILITY STATEMENT

535 We are committed to ensuring the reproducibility of our results. All datasets used in this paper are 536 publicly available, and detailed descriptions of preprocessing steps are provided in the appendix. 537 In addition, we will release our source code, trained models, and experiment configurations upon 538 publication to facilitate replication of our experiments. This includes scripts for data preprocessing, 539 training, and evaluation, ensuring that other researchers can reproduce our findings.
540

540 REFERENCES
541

542 Yongqi Fan, Xiaoyang Chen, Dezhi Ye, Jie Liu, Haijin Liang, Jin Ma, Ben He, Yingfei Sun, and
543 Tong Ruan. Tfrank: Think-free reasoning enables practical pointwise llm ranking, 2025. URL
544 <https://arxiv.org/abs/2508.09539>.

545 Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural
546 networks, 2017. URL <https://arxiv.org/abs/1706.04599>.

547 Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu,
548 Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms
549 via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

550 Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec
551 Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv
552 preprint arXiv:2412.16720*, 2024.

553 Nour Jedidi, Yung-Sung Chuang, James Glass, and Jimmy Lin. Don't "overthink" passage rerank-
554 ing: Is reasoning truly necessary?, 2025. URL <https://arxiv.org/abs/2505.16886>.

555 Yuelyu Ji, Zhuochun Li, Rui Meng, and Daqing He. Reasoningrank: Teaching student models
556 to rank through reasoning-based knowledge distillation, 2025. URL [https://arxiv.org/
557 abs/2410.05168](https://arxiv.org/abs/2410.05168).

558 Dohyeon Lee, Yeonseok Jeong, and Seung won Hwang. From token to action: State machine
559 reasoning to mitigate overthinking in information retrieval, 2025. URL [https://arxiv.org/
560 abs/2505.23059](https://arxiv.org/abs/2505.23059).

561 Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal,
562 Heinrich Kütller, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented gener-
563 ation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:
564 9459–9474, 2020.

565 Wenhan Liu, Xinyu Ma, Weiwei Sun, Yutao Zhu, Yuchen Li, Dawei Yin, and Zhicheng Dou.
566 Reasonrank: Empowering passage ranking with strong reasoning ability, 2025. URL <https://arxiv.org/abs/2508.07050>.

567 Xuan Lu, Sifan Liu, Bochao Yin, Yongqi Li, Xinghao Chen, Hui Su, Yaohui Jin, Wenjun Zeng, and
568 Xiaoyu Shen. Multiconir: Towards multi-condition information retrieval, 2025. URL <https://arxiv.org/abs/2503.08046>.

569 Xueguang Ma, Liang Wang, Nan Yang, Furu Wei, and Jimmy Lin. Fine-tuning llama for multi-stage
570 text retrieval. In *Proceedings of the 47th International ACM SIGIR Conference on Research and
571 Development in Information Retrieval*, pp. 2421–2425, 2024.

572 Tong Niu, Shafiq Joty, Ye Liu, Caiming Xiong, Yingbo Zhou, and Semih Yavuz. Judgerank:
573 Leveraging large language models for reasoning-intensive reranking, 2024. URL <https://arxiv.org/abs/2411.00142>.

574 Rodrigo Nogueira, Wei Yang, Kyunghyun Cho, and Jimmy Lin. Multi-stage document ranking with
575 bert, 2019. URL <https://arxiv.org/abs/1910.14424>.

576 Rodrigo Frassetto Nogueira, Zhiying Jiang, Ronak Pradeep, and Jimmy Lin. Document ranking
577 with a pretrained sequence-to-sequence model. In *EMNLP (Findings)*, 2020.

578 Ronak Pradeep, Sahel Sharifmoghaddam, and Jimmy Lin. Rankvicuna: Zero-shot listwise docu-
579 ment reranking with open-source large language models, 2023a. URL [https://arxiv.org/
580 abs/2309.15088](https://arxiv.org/abs/2309.15088).

581 Ronak Pradeep, Sahel Sharifmoghaddam, and Jimmy Lin. Rankzephyr: Effective and robust zero-
582 shot listwise reranking is a breeze!, 2023b. URL <https://arxiv.org/abs/2312.02724>.

594 Zhen Qin, Rolf Jagerman, Kai Hui, Honglei Zhuang, Junru Wu, Le Yan, Jiaming Shen, Tianqi
 595 Liu, Jialu Liu, Donald Metzler, Xuanhui Wang, and Michael Bendersky. Large language mod-
 596 els are effective text rankers with pairwise ranking prompting. In Kevin Duh, Helena Gomez,
 597 and Steven Bethard (eds.), *Findings of the Association for Computational Linguistics: NAACL*
 598 2024, pp. 1504–1518, Mexico City, Mexico, June 2024. Association for Computational Lin-
 599 guistics. doi: 10.18653/v1/2024.findings-naacl.97. URL <https://aclanthology.org/2024.findings-naacl.97/>.

600

601 Yuxin Ren, Qiya Yang, Yichun Wu, Wei Xu, Yalong Wang, and Zhiqiang Zhang. Non-autoregressive
 602 generative models for reranking recommendation. In *Proceedings of the 30th ACM SIGKDD*
 603 Conference on Knowledge Discovery and Data Mining

604 Conference on Knowledge Discovery and Data Mining, pp. 5625–5634, 2024.

605 Rulin Shao, Rui Qiao, Varsha Kishore, Niklas Muennighoff, Xi Victoria Lin, Daniela Rus, Bryan
 606 Kian Hsiang Low, Sewon Min, Wen tau Yih, Pang Wei Koh, and Luke Zettlemoyer. Reasonir:
 607 Training retrievers for reasoning tasks, 2025. URL <https://arxiv.org/abs/2504.20595>.

608

609 Sahel Sharifmoghaddam, Ronak Pradeep, Andre Slavescu, Ryan Nguyen, Andrew Xu, Zijian
 610 Chen, Yilin Zhang, Yidi Chen, Jasper Xian, and Jimmy Lin. Rankilm: A python package for
 611 reranking with llms. In *Proceedings of the 48th International ACM SIGIR Conference on Re-*
 612 *search and Development in Information Retrieval*, pp. 3681–3690, 2025.

613

614 Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin,
 615 and Zhaochun Ren. Is chatgpt good at search? investigating large language models as re-ranking
 616 agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Pro-*
 617 *cessing*, pp. 14918–14937, 2023a.

618 Weiwei Sun, Lingyong Yan, Xinyu Ma, Shuaiqiang Wang, Pengjie Ren, Zhumin Chen, Dawei Yin,
 619 and Zhaochun Ren. Is chatgpt good at search? investigating large language models as re-ranking
 620 agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Pro-*
 621 *cessing*, pp. 14918–14937, 2023b.

622 William Webber, Alistair Moffat, and Justin Zobel. A similarity measure for indefinite rankings.
 623 *ACM Transactions on Information Systems (TOIS)*, 28(4):1–38, 2010.

624

625 Orion Weller, Kathryn Ricci, Eugene Yang, Andrew Yates, Dawn Lawrie, and Benjamin Van
 626 Durme. Rank1: Test-time compute for reranking in information retrieval, 2025. URL <https://arxiv.org/abs/2502.18418>.

627

628 Eugene Yang, Andrew Yates, Kathryn Ricci, Orion Weller, Vivek Chari, Benjamin Van Durme,
 629 and Dawn Lawrie. Rank-k: Test-time reasoning for listwise reranking, 2025. URL <https://arxiv.org/abs/2505.14432>.

630

631 Le Zhang, Bo Wang, Xipeng Qiu, Siva Reddy, and Aishwarya Agrawal. Rearank: Reasoning re-
 632 ranking agent via reinforcement learning, 2025. URL <https://arxiv.org/abs/2505.20046>.

633

634 Shengyao Zhuang, Honglei Zhuang, Bevan Koopman, and Guido Zuccon. A setwise approach
 635 for effective and highly efficient zero-shot ranking with large language models. In *Proceedings*
 636 *of the 47th International ACM SIGIR Conference on Research and Development in Information*
 637 *Retrieval*, pp. 38–47, 2024.

638

639 Shengyao Zhuang, Xueguang Ma, Bevan Koopman, Jimmy Lin, and Guido Zuccon. Rank-r1:
 640 Enhancing reasoning in llm-based document rerankers via reinforcement learning, 2025. URL
 641 <https://arxiv.org/abs/2503.06034>.

642

643

644

645

646

647

648 **A BACKBONES OF RERANKERS**
649650 Table 5 summarizes the backbones and training strategies of both existing reasoning-enhanced
651 rerankers and our proposed models. We observe that the **Qwen** family has become the mainstream
652 backbone for LLM-based reranking. Our proposed *Direct Rerankers* and *Reason Rerankers* are
653 built upon Qwen3-4B and Qwen3-8B, ensuring a fair comparison with recent reasoning-enhanced
654 baselines while highlighting the impact of reasoning versus direct decision-making.
655656 Table 5: Overview of baseline and proposed rerankers.
657

658 Model	659 Training	660 Backbone	661 Type
660 BM25	661 /	662 /	663 Pointwise
661 Rank1-7B	662 SFT	663 Qwen2.5-7B	664 Pointwise
662 Rank1-14B	663 SFT	664 Qwen2.5-14B	665 Pointwise
663 TFRank-4B	664 SFT + GRPO	665 Qwen3-4B	666 Pointwise
664 TFRank-8B	665 SFT + GRPO	666 Qwen3-8B	667 Pointwise
665 REARANK-7B	666 GRPO	667 Qwen2.5-7B	668 Listwise
666 Rank-R1-7B	667 GRPO	668 Qwen2.5-7B	669 Listwise
667 Rank-R1-14B	669 GRPO	670 Qwen2.5-14B	671 Listwise
668 ReasonRank-7B	671 SFT + GRPO	672 Qwen2.5-7B	673 Listwise
669 Direct-Point-4B	673 SFT	674 Qwen3-4B	675 Pointwise
670 Direct-Point-8B	675 SFT	676 Qwen3-8B	677 Pointwise
671 Reason-Point-4B	677 SFT	678 Qwen3-4B	679 Pointwise
672 Reason-Point-8B	680 SFT	681 Qwen3-8B	682 Pointwise
673 Direct-List-4B	683 SFT + GRPO	684 Qwen3-4B	685 Listwise
674 Direct-List-8B	685 SFT + GRPO	686 Qwen3-8B	687 Listwise
675 Reason-List-4B	687 SFT + GRPO	688 Qwen3-4B	689 Listwise
676 Reason-List-8B	689 SFT + GRPO	690 Qwen3-8B	691 Listwise

677 **B EXPERIMENTAL RESULTS FOR GPT4_REASON QUERY**
678679 Table 6 reports the performance of Reason rerankers and Direct rerankers under different training
680 stages on BRIGHT with gpt4_reason queries. The results are consistent with the findings in
681 Section 3.4, showing that non-CoT Direct rerankers consistently outperform their reasoning
682 counterparts. Table 7 further presents the performance of Direct rerankers on BRIGHT with gpt4_reason
683 queries. Among pointwise models, Direct-Point-4B achieves the best score of 33.3, followed by
684 Direct-Point-8B with 32.0. For listwise models, both Direct-List-4B and Direct-List-8B obtain 35.3,
685 outperforming reasoning-enhanced rerankers and providing further evidence that explicit reasoning
686 does not lead to better reranking performance.
687688 **C PROMPTS FOR RERANKING**
689690 **C.1 PROMPT FOR POINTWISE RERANKING**
691692 In the pointwise setting, the reranker judges each query–passage pair independently. The non-
693 reasoning version (**Direct-Point**) directly outputs a binary decision, as shown in Figure 5. The
694 reasoning version (**Reason-Point**) additionally generates a rationale enclosed within `<think>` tags
695 before giving the final decision, as shown in Figure 4.
696697 **C.2 PROMPT FOR LISTWISE RERANKING**
698699 In the listwise setting, the reranker considers the entire candidate set and outputs a ranked order
700 of passages. The non-reasoning version (**Direct-List**) directly produces the final ranking sequence,
701 as shown in Figure 7. The reasoning version (**Reason-List**) first generates a reasoning trace in
702 `<think>` tags and then outputs the final ranking within `<answer>` tags, as shown in Figure 6.

702
703
704
705 Table 6: Performance of Direct-Point, Reason-Point, Direct-List, and Reason-List under different
706 training strategies on BRIGHT (gpt4-query).

705 706 Model	707 Training	708 StackExchange						709 Coding		710 Theorem-based			711 Avg.	
		712 Bio.	713 Earth.	714 Econ.	715 Psy.	716 Rob.	717 Stack.	718 Sus.	719 Leet.	720 Pony	721 AoPS	722 TheoQ.		
Pointwise														
Reason-Point-4B	SFT	46.2	49.8	26.1	37.0	22.0	26.2	28.4	24.7	23.9	6.5	33.5	20.2	28.7
Direct-Point-4B	SFT	50.8	54.7	31.2	41.7	25.8	30.6	32.1	29.1	28.8	7.8	38.5	25.0	33.0
Reason-Point-8B	SFT	42.0	44.5	27.0	36.4	23.1	28.7	30.2	28.1	18.9	7.5	24.5	19.5	27.5
Direct-Point-8B	SFT	46.9	49.2	31.5	41.7	27.0	33.6	35.3	34.6	22.5	9.7	28.8	22.9	<u>32.0</u>
Listwise														
Reason-List-4B	SFT	50.9	45.5	27.5	38.0	28.5	28.8	34.0	20.1	21.5	6.9	24.0	31.0	29.7
Direct-List-4B	SFT	53.5	48.3	29.8	39.9	31.2	30.5	36.9	23.0	25.0	8.2	26.3	34.2	32.2
Reason-List-8B	SFT	51.4	47.0	29.5	40.1	29.0	29.9	35.5	21.9	26.2	7.7	27.5	32.0	31.5
Direct-List-8B	SFT	54.0	49.6	30.1	42.0	31.6	31.8	37.1	24.1	28.0	8.4	28.1	35.0	33.3
Reason-List-4B	SFT+GRPO	55.1	49.0	28.7	42.7	31.1	33.1	36.6	21.8	23.8	7.8	27.7	37.2	32.9
Direct-List-4B	SFT+GRPO	58.4	51.8	31.3	41.6	34.4	33.4	41.0	25.2	27.9	9.8	29.5	39.2	35.3
Reason-List-8B	SFT+GRPO	54.5	49.4	30.8	44.4	29.9	32.1	38.7	22.6	28.0	8.9	31.4	36.0	33.9
Direct-List-8B	SFT+GRPO	56.1	52.2	30.3	44.8	32.5	36.4	38.8	25.0	30.9	8.5	29.6	38.9	<u>35.3</u>

721
722 Table 7: Performance of different rerankers on BRIGHT datasets (gpt4_query).

723 Model	724 Training	725 StackExchange						726 Coding		727 Theorem-based			728 Avg.	
		729 Bio.	730 Earth.	731 Econ.	732 Psy.	733 Rob.	734 Stack.	735 Sus.	736 Leet.	737 Pony	738 AoPS	739 TheoQ.	TheoT.	
Pointwise														
BM25	/	53.6	53.6	24.3	38.6	18.8	22.7	25.9	17.7	19.3	3.9	20.2	18.9	26.5
MonoT5-3B	SFT	16.0	24.0	17.7	19.5	8.0	10.5	19.5	17.2	29.2	7.1	20.3	12.0	16.8
RankLLaMA-7B	SFT	17.5	15.5	13.1	13.6	17.9	6.9	16.9	8.4	46.8	2.2	4.5	3.5	13.9
RankLLaMA-13B	SFT	21.6	19.1	16.3	14.0	15.7	7.7	18.5	8.8	31.1	1.7	4.4	4.9	13.7
Rank1-7B	SFT	48.8	36.7	20.8	35.0	22.0	18.7	36.2	12.7	31.2	6.3	23.7	37.8	27.5
Rank1-14B	SFT	49.3	37.7	22.6	35.2	22.5	20.8	33.6	17.7	33.2	8.4	22.5	41.4	28.7
Rank1-32B	SFT	49.7	35.8	22.0	37.5	22.5	21.7	35.0	18.8	32.5	10.8	22.9	43.7	29.4
Direct-Point-4B	SFT	50.8	54.7	31.2	41.7	25.8	30.6	32.1	29.1	28.8	7.8	38.5	25.0	33.0
Direct-Point-8B	SFT	46.9	49.2	31.5	41.7	27.0	33.6	35.3	34.6	22.5	9.7	28.8	22.9	<u>32.0</u>
Listwise														
RankZephyr-7B	SFT	44.1	31.0	17.9	28.4	17.5	27.0	21.6	18.9	17.8	2.7	15.9	12.7	21.3
Rank-K	SFT	50.4	46.2	30.6	46.7	32.4	33.0	41.2	24.0	32.2	7.6	28.3	26.6	<u>33.3</u>
ReasonRank-7B	SFT+GRPO	56.4	51.2	28.4	43.4	31.0	31.9	39.1	23.0	7.6	8.1	29.9	39.1	32.4
Direct-List-4B	SFT+GRPO	58.4	51.8	31.3	41.6	34.4	33.4	41.0	25.2	27.9	9.8	29.5	39.2	35.3
Direct-List-8B	SFT+GRPO	56.1	52.2	30.3	44.8	32.5	36.4	38.8	25.0	30.9	8.5	29.6	38.9	35.3

741
742

D EXAMPLES FOR RANKING OUTPUT

743
744

D.1 EXAMPLES FOR POINTWISE RANKING

745
746 Direct-Point rerankers perform a single forward pass per input and predict a binary answer token
747 (true/false); we report the logits at the answer position, so the model does not explicitly print
748 the answer string. In contrast, Reason-Point first generates a natural-language rationale and then
749 computes the logits at the answer position, thereby outputting the complete reasoning text. For
750 clarity, Figure 8 presents an output instance, where we additionally decode the answer-position
751 logits into the corresponding answer token.752
753

D.2 EXAMPLES FOR LISTWISE RANKING

754
755 Figures 6 and 10 illustrate a concrete example of the listwise reranking setting. Figure 6 shows the
input prompt template provided to the model, where the query and candidate passages are listed, and
the model is instructed to return a complete ranking. Figure 10 presents the corresponding outputs

```

756
757 <|im_start|>system
758 Determine if the following passage is relevant
759 to the query. Answer only with 'true' or 'false'.
760 <|im_end|>
761 <|im_start|>user
762 Query: {}
763 Passage: {}
764 <|im_end|>
765 <|im_start|>assistant
766

```

Figure 4: Prompt template for pointwise relevance judgement.

```

769
770 <|im_start|>system
771 Determine if the following passage is relevant
772 to the query. Answer only with 'true' or 'false'.
773 <|im_end|>
774 <|im_start|>user
775 Query: {}
776 Passage: {}
777 <|im_end|>
778 <|im_start|>assistant
779 <think> </think>
780

```

Figure 5: Prompt template for pointwise relevance judgement (non-reasoning version). The <think> tag is kept empty to maintain consistency with reasoning prompts.

from different variants of our models. Non-reasoning models (Direct-List) directly produce the ranked sequence enclosed in <answer> tags, while reasoning models (Reason-List) first generate an explicit rationale enclosed in <think> tags before outputting the final ranking. This comparison highlights how reasoning influences the ranking process, providing intermediate explanations at the cost of increased verbosity.

E ADDITIONAL GENERALIZATION EXPERIMENTS

To address the reviewer’s request for verification beyond the Qwen3 backbone and the inclusion of zero-shot baselines, we conducted two additional sets of experiments: (1) Generalization across model families. We trained **Qwen2.5-7B** and **Llama-3.2-8B** rerankers using the same pointwise and listwise configurations as reported in the main paper. (2) Zero-shot baselines. We evaluated **Qwen3-8B** in zero-shot mode under both pointwise and listwise settings.

As shown in Figure 8, across all examined backbones—**Qwen3**, **Qwen2.5**, and **Llama-3.2**—we observe a highly consistent pattern: *explicit reasoning does not yield improvements under either pointwise or listwise reranking*. This confirms that our findings are architecture-agnostic rather than specific to a single model family. These results strengthen the generality of our conclusions and confirm that the observed reasoning-induced degradation is not tied to a specific model family.

F ANALYSIS OF PROMPT SENSITIVITY AND TRAINING STABILITY

This appendix provides additional analyses to examine whether the performance gap between Direct and Reason rerankers may be attributed to prompt sensitivity, checkpoint instability, or randomness from training seeds. The results indicate that the observed performance differences are systematic rather than arising from experimental artifacts.

```

810
811 You are RankLLM, an intelligent assistant that can
812 rank passages based on their relevance to the query.
813 Given a query and a passage list, you first think
814 about the reasoning process in the mind and then
815 provide the answer (i.e., the reranked passage list).
816
817 The reasoning process and answer are enclosed
818 within <think> </think> and <answer> </answer>
819 tags, respectively, i.e.,
820 <think> reasoning process here </think>
821 <answer> answer here </answer>.
822
823 I will provide you with {num} passages, each
824 indicated by a numerical identifier [].
825 Rank the passages based on their relevance to
826 the search query:
827
828 [1]: {{passage_1}}
829 [2]: {{passage_2}}
830 (more passages)...
831
832 Search Query: {{query}}.
833 Rank the {num} passages above based on their
834 relevance to the search query.
835
836 All passages should be included and listed using
837 identifiers, in descending order of relevance.
838 The format of the answer should be [] > [],
839 e.g., [2] > [1].
840
841
842

```

Figure 6: Prompt template for listwise reranking with explicit reasoning. The model produces a reasoning trace within `<think>` tags and a final ranking within `<answer>` tags.

Table 8: Reranking performance across different model families. Metrics are NDCG@10 on BRIGHT and BEIR.

Model Variant	Backbone	BRIGHT	BEIR
Direct-Point	Qwen2.5-7B	21.6	41.3
Reason-Point	Qwen2.5-7B	18.3	39.4
Direct-List	Qwen2.5-7B	22.3	40.6
Reason-List	Qwen2.5-7B	20.5	38.3
Direct-Point	Llama-3.2-8B	22.8	42.1
Reason-Point	Llama-3.2-8B	19.0	39.7
Direct-List	Llama-3.2-8B	23.7	41.0
Reason-List	Llama-3.2-8B	22.8	39.8
Direct-Point	Qwen3-8B (zero-shot)	16.9	38.4
Reason-Point	Qwen3-8B (zero-shot)	15.4	37.3
Direct-List	Qwen3-8B (zero-shot)	17.3	39.2
Reason-List	Qwen3-8B (zero-shot)	16.4	37.8

F.1 PROMPT SENSITIVITY EVALUATION

To assess whether the gap is caused by instruction phrasing rather than model behavior, we conducted a semantic prompt-swap study on both BRIGHT and BEIR. The original Rank1-style instruction was replaced with (i) a paraphrased semantic rewrite and (ii) a minimal instruction with no dataset hint, while preserving task semantics. As shown in Table 9, across all variants, the relative ranking between Direct and Reason models remains stable, indicating that the performance gap does not arise from prompt sensitivity.

```

864
865 You are RankLLM, an intelligent assistant that can
866 rank passages based on their relevance to the query.
867 Given a query and a passage list, directly provide
868 the reranked passage list without generating any
869 reasoning process.
870
871 I will provide you with {num} passages, each
872 indicated by a numerical identifier [].
873 Rank the passages based on their relevance to
874 the search query:
875 [1]: {{passage_1}}
876 [2]: {{passage_2}}
877 (more passages) ...
878
879 Search Query: {{query}}.
880 Rank the {num} passages above based on their
881 relevance to the search query.
882
883 All passages should be included and listed using
884 identifiers, in descending order of relevance.
885 The format of the answer should be [] > [],
886 e.g., [2] > [1].
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2197
2198
2199
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2238
2239
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2248
2249
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2258
2259
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2268
2269
2269
2270
2271
2272
2273
2274
2275
2276
2277
2277
2278
2278
2279
2279
2280
2281
2282
2283
2284
2285
2286
2287
2287
2288
2288
2289
2289
2290
2291
2292
2293
2294
2295
2296
2297
2297
2298
2298
2299
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2328
2329
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2338
2339
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2348
2349
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2368
2369
2369
2370
2371
2372
2373
2374
2375
2376
2377
2377
2378
2378
2379
2379
2380
2381
2382
2383
2384
2385
2386
2387
2387
2388
2388
2389
2389
2390
2391
2392
2393
2394
2395
2396
2397
2397
2398
2398
2399
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2418
2419
2419
2420
2421
2422
2423
2424
2425
2426
2427
2427
2428
2428
2429
2429
2430
2431
2432
2433
2434
2435
2436
2437
2437
2438
2438
2439
2439
2440
2441
2442
2443
2444
2445
2446
2447
2447
2448
2448
2449
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2458
2459
2459
2460
2461
2462
2463
2464
2465
2466
2467
2467
2468
2468
2469
2469
2470
2471
2472
2473
2474
2475
2476
2477
2477
2478
2478
2479
2479
2480
2481
2482
2483
2484
2485
2486
2487
2487
2488
2488
2489
2489
2490
2491
2492
2493
2494
2495
2496
2497
2497
2498
2498
2499
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2518
2519
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2528
2529
2529
2530
2531
2532
2533
2534
2535
2536
2537
2537
2538
2538
2539
2539
2540
2541
2542
2543
2544
2545
2546
2547
2547
2548
2548
2549
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2558
2559
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2568
2569
2569
2570
2571
2572
2573
2574
2575
2576
2577
2577
2578
2578
2579
2579
2580
2581
2582
2583
2584
2585
2586
2587
2587
2588
2588
2589
2589
2590
2591
2592
2593
2594
2595
2596
2597
2597
2598
2598
2599
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2618
2619
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2628
2629
2629
2630
2631
2632
2633
2634
2635
2636
2637
2637
2638
2638
2639
2639
2640
2641
2642
2643
2644
2645
2646
2647
2647
2648
2648
2649
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2658
2659
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2668
2669
2669
2670
2671
2672
2673
2674
2675
2676
2677
2677
2678
2678
2679
2679
2680
2681
2682
2683
2684
2685
2686
2687
2687
2688
2688
2689
2689
2690
2691
2692
2693
2694
2695
2696
2697
2697
2698
2698
2699
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2718
2719
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2728
2729
2729
2730
2731
2732
2733
2734
2735
2736
2737
2737
2738
2738
2739
2739
2740
2741
2742
2743
2744
2745
2746
2747
2747
2748
2748
2749
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2758
2759
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2768
2769
2769
2770
2771
2772
2773
2774
2775
2776
2777
2777
2778
2778
2779
2779
2780
2781
2782
2783
2784
2785
2786
2787
2787
2788
2788
2789
2789
2790
2791
2792
2793
2794
2795
2796
2797
2797
2798
2798
2799
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2809
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2818
2819
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2828
2829
2829
2830
2831
2832
2833
2834
2835
2836
2837
2837
2838
2838
2839
2839
2840
2841
2842
2843
2844
2845
2846
2847
2847
2848
2848
2849
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2858
2859
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2868
2869
2869
2870
2871
2872
2873
2874
2875
2876
2877
2877
2878
2878
2879
2879
2880
2881
2882
2883
2884
2885
2886
2887
2887
2888
2888
2889
2889
2890
2891
2892
2893
2894
2895
2896
2897
2898
2899
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2909
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2918
2919
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2928
2929
2929
2930
2931
2932
2933
2934
2935
2936
2937
2937
2938
2938
2939
2939
2940
2941
2942
2943
2944
2945
2
```

```

918
919 <|im_start|>system
920 Determine if the following passage is relevant
921 to the query. Answer only with 'true' or 'false'.
922 <|im_end|>
923 <|im_start|>user
924 Query: Have governments ever defaulted on their
925 domestic public debt or treasury bonds due to
926 hyperinflation or political crises?
927 Passage: Argentina experienced a unique default
928 when it suspended all debt payments before a legal
929 default was declared, with 60% of the debt held
930 by Argentines. Rating agencies still classified it
931 as a default. Russia defaulted on its domestic debt
932 (GKOs) in 1998, which became one of the most
933 significant financial crises of that decade.
934 <|im_end|>
935 <|im_start|>assistant
936 <think>
937 Okay, let's tackle this query. The user is asking if
938 governments have ever defaulted on their public domestic
939 debt or treasury bonds, specifically due to hyperinflation
940 or political issues. The passage mentions Argentina and
941 Russia. Argentina suspended all payments, considered a
942 default by agencies, and Russia defaulted on GKOs in 1998.
943 Both are clear cases of default on domestic debt. While the
944 passage does not explicitly state hyperinflation, these are
945 examples that confirm the query.
946 ...
947 </think>
948 true
949
950
951
952
953
954

```

Figure 8: Example of a Reason-Pointwise output with explicit reasoning in the `<think>` block before producing the binary decision.

Seed	Direct-List-8B (BRIGHT)	Reason-List-8B (BRIGHT)
42 (default)	27.1	25.9
0	26.9	25.7

Table 11: Random seed robustness evaluation. Results remain consistent across training runs.

G REASONING ASSUMPTION IN RANKING AND VERIFICATION OF REASONING TRACE QUALITY

This section presents quantitative and qualitative analyses to verify that the reasoning-based models generate valid and meaningful reasoning traces. The goal is to ensure that the observed performance trends are not attributable to poorly trained or defective reasoning behavior.

G.1 REASONING ASSUMPTION IN RANKING

Reranking with explicit reasoning implicitly relies on a core assumption: *the generated chain-of-thought should provide a logically valid, coherent, and decision-supportive explanation that reflects the model's underlying relevance judgment.*

This assumption has been adopted—either explicitly or implicitly—in recent reasoning-enhanced ranking systems, where the reasoning trace is expected to: (1) extract or reference query-document

```

972
973 Search Query: where is a elephant habitat.
974
975 Passages:
976 [1] Elephant Habitat. Elephants make home in a variety of habitats
977 including tropical and subtropical zones...
978 [2] Elephant Natural Habitat. Elephants are able to survive in a
979 variety of different locations...
980 [3] This is because the elephant is well known to being intelligent
981 and capable of experiencing...
982 ...
983 [20] Elephant Habitat. Elephants make home in a variety of habitats
984 including tropical and subtropical zones...
985
986 Rank the 20 passages above based on their relevance to the query.
987 All the passages should be included and listed using identifiers, in
988 descending order.
989 The format of the answer should be [ ] > [ ], e.g., [2] > [1].
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025

```

Figure 9: Example listwise prompt for the query “*where is a elephant habitat*” with 20 candidate passages.

evidence relevant to the final decision; (2) maintain internal logical consistency; and (3) articulate the decision boundary in a manner aligned with the teacher or supervision signal.

Under this assumption, if explicit reasoning improves the quality of the relevance decision, we would expect the reasoning-based reranker to produce *accurate, coherent, and faithful reasoning traces*, which subsequently translate into better ranking performance.

Therefore, our analyses proceed from the following question: *Do reasoning-based rerankers indeed learn to generate high-quality reasoning traces that support the relevance decision?*

To answer this, we conduct both quantitative and qualitative verification of reasoning trace quality in the following sections.

G.2 QUANTITATIVE VERIFICATION

To assess whether the reasoning model has successfully learned to produce teacher-level decisions, we evaluate its binary true/false classification accuracy without applying any ranking procedure. The reported results (Table 3) compare the teacher model (DEEPSEEK-R1), the reasoning-based student (REASON-POINT-8B), and the direct baseline (DIRECT-POINT-8B) across both in-domain (MS MARCO) and out-of-domain (BRIGHT-Biology) settings.

The REASON-POINT-8B model achieves an average accuracy of **67.5**, closely matching the teacher model (**68.6**). This demonstrates that the student has successfully learned the teacher’s decision behavior, indicating that the performance degradation observed in ranking cannot be attributed to failed reasoning acquisition.

G.3 QUALITATIVE VERIFICATION

We further assess the validity of generated reasoning traces through a targeted qualitative study. A total of 50 query–document pairs were randomly sampled from MS MARCO and BRIGHT. For each pair, reasoning chains produced by DEEPSEEK-R1 and REASON-POINT-8B were independently evaluated by three annotators using a three-dimensional rubric:

- **Correctness:** logical appropriateness of the conclusion,
- **Coherence:** internal consistency of reasoning steps,
- **Faithfulness:** structural alignment with the teacher’s reasoning.

```

1026
1027 Direct-List-8B-SFT:
1028 <think>
1029 </think>
1030 <answer> [8] > [19] > [11] > ... > [14] > [3] > [4] </answer>
1031
1032 Reason-List-8B-SFT:
1033 <think>
1034 Okay, let's tackle this ranking problem. The user wants to know where
1035 elephants live.
1036 I need to focus on passages that explicitly describe habitats (
1037 savannas, forests, deserts, etc.).
1038 Passage [1] lists tropical zones, dry forests, savannas, semi-deserts,
1039 rainforests.
1040 [2] and [5] mention grasslands, deserts, forests, swamps.
1041 [8] mentions Africa and South Asia. [10] and [11] describe
1042 adaptability.
1043 Passages [4], [6], [7], [9], [19] repeat similar information.
1044 [13] talks about different species. [15] and [17] focus on
1045 conservation.
1046 [18] and [20] similar to [1] and [16].
1047 Less direct: [3], [14], [12].
1048 So top: [1], [2], [5], [8], [10], [11], [19], [13], [16], [20].
1049 </think>
1050 <answer> [1] > [2] > [5] > ... > [12] > [3] > [14] </answer>
1051
1052 Direct-List-8B-GRPO:
1053 <think>
1054 </think>
1055 <answer> [8] > [19] > [1] > ... > [6] > [13] > [14] </answer>
1056
1057 Reason-List-8B-GRPO:
1058 <think>
1059 Okay, let's tackle this query about where elephants live. The user
1060 wants to know their habitats.
1061 Passages that directly mention savannas, forests, deserts, and
1062 rainforests are key.
1063 Looking at [1], [2], [5], [8], [10], [11], [16], [19], [20] all list
1064 habitats.
1065 [8] and [19] mention Africa and South Asia, which is important.
1066 Others talk about captivity or human impact, less relevant.
1067 So the top ones are those listing habitats explicitly.
1068 </think>
1069 <answer> [1] > [2] > [5] > ... > [17] > [18] > [14] </answer>
1070
1071

```

Figure 10: Examples of model outputs on the elephant habitat ranking task. Non-reasoning models (NL) directly output rankings, while reasoning models (RL) generate intermediate rationales within <think> tags before the final answer.

Majority-vote results are summarized in Table 12.

Evaluation Dimension	DeepSeek-R1	Reason-Point-8B	Agreement (%)
Correctness	88%	86%	83%
Coherence	92%	88%	81%
Faithfulness	—	83%	—

Table 12: Qualitative evaluation of reasoning trace quality.

1080
1081 The results show that REASON-POINT-8B produces reasoning traces that are largely correct (86%),
1082 coherent (88%), and structurally aligned with the teacher (83%). These findings confirm that the
1083 model generates valid and well-formed reasoning, and that the observed ranking behavior is not
1084 attributable to defective or low-quality reasoning traces.
1085

1086 H THE USE OF LARGE LANGUAGE MODELS (LLMs)

1087 In the preparation of this manuscript, we employed GPT-5 to assist with text refinement. Specifi-
1088 cally, the model was used to improve the clarity, readability, and overall presentation of the paper
1089 by correcting grammatical errors, smoothing sentence structures, and enhancing stylistic consis-
1090 tency. Importantly, all core research ideas, experimental designs, and results were conceived and
1091 validated by the authors; the role of the LLM was limited to linguistic polishing. This ensured that
1092 the scientific content remained entirely authored by the researchers, while benefiting from improved
1093 academic writing quality.
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133