Lightweight and Direct Document Relevance Optimization for Generative Information Retrieval

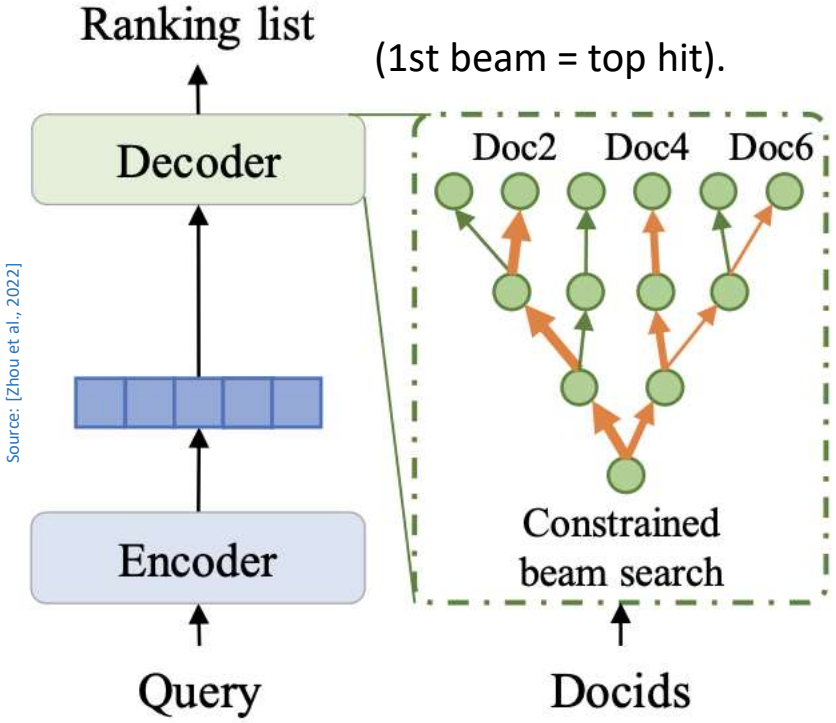
Kidist Amde Mekonnen **Yubao Tang** **Maarten de Rijke**

IR Lab AMS

UNIVERSITY OF AMSTERDAM

SIGIR 2025 Padova ITALY

Introduction → Generative / Autoregressive Retrieval



Reframing Search

- Index = Model training
- Retrieval = Generation
- End-to-End differentiable

➢ A prefix tree constrains generation to valid outputs within the target space.

Motivation → Misalignment in Generative IR (GenIR)

- **Learning Gap:** Optimized to spell doc IDs, not rank by relevance.
- Fluency ≠ Relevance: Next-token likelihood rewards fluency, not retrieval quality.
- Ranking Drift: Beam order mirrors token probability, not true relevance.
- Objective Mismatch: Token-level cross-entropy ≠ document-level ranking signal.

➢ **Motivation**: close the gap with a single end-to-end objective that links generation probabilities directly to document relevance.

Related Work & Positioning

Prior approaches address misalignment, but rely on heavy, indirect solutions.

- **GenRRL** *(EMNLP '23)*: Uses reward models and policy-gradient RL.
- **LTRGR** *(AAAI '24)*: Adds list-wise margin loss post-hoc via heuristic mappings.
- **ROGER** *(TOIS '24)*: Distils relevance from dense retrievers.

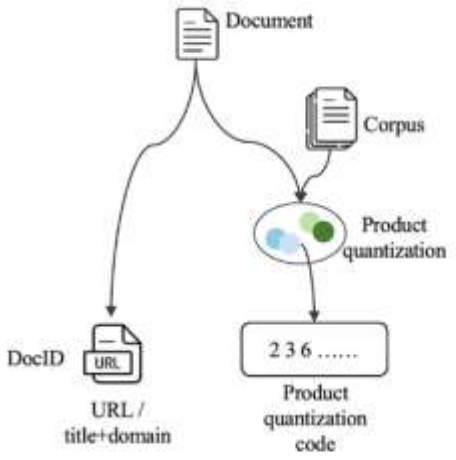
➢ **Gap**: we still lack a *direct, lightweight* way to make the generator's own log-probs align with document relevance.

Our Goal → From Heavy Pipelines to Lightweight DDRO

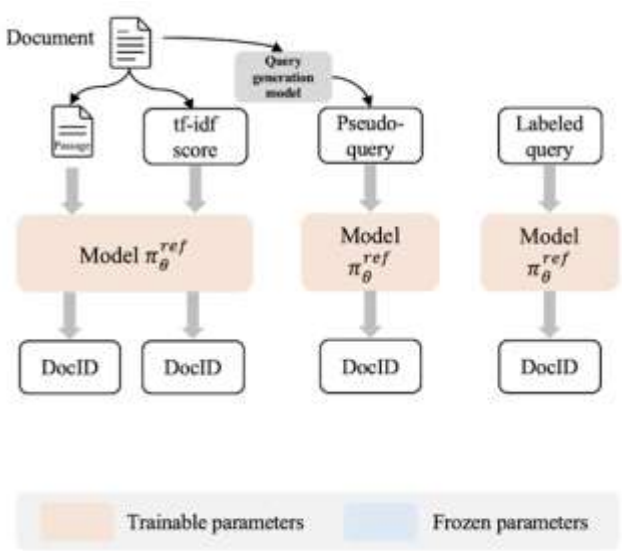
- Retain alignment without RL, dense-teacher distillation, or heuristic re-ranking.
- Optimize generative retrieval using a **KL-regularized pairwise objective.**
- Replace complex components with a simple, effective learning-to-rank step.
- Inference: Same beam decoding; scores now reflect relevance.

Workflow Overview: Three Core Steps

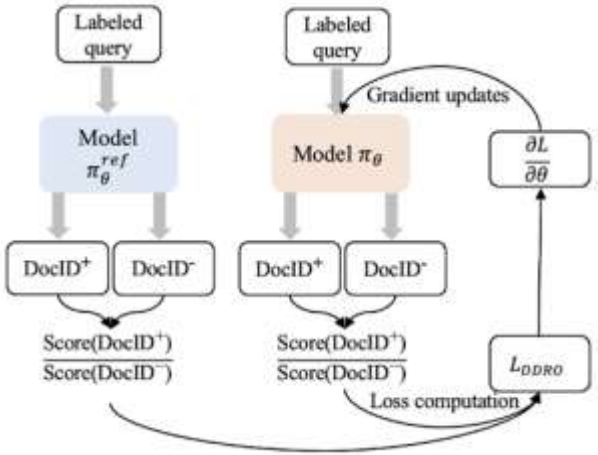
Step 1: DocID construction



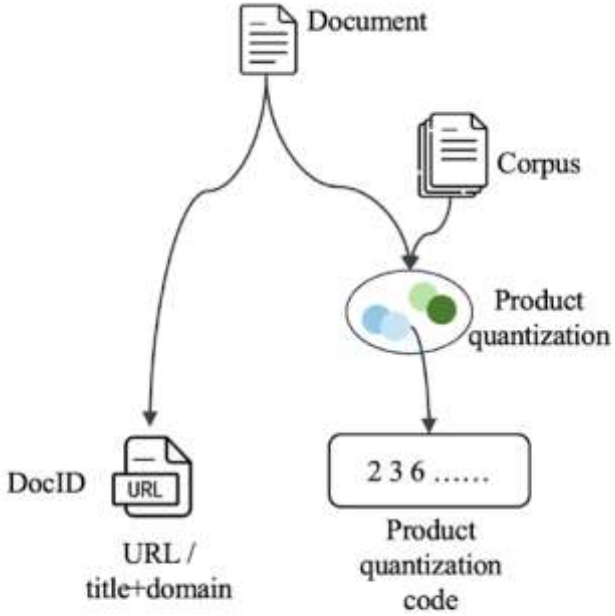
Step 2: Supervised fine-tuning

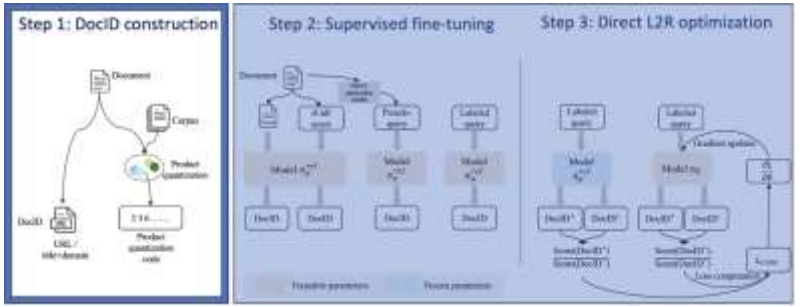


Step 3: Direct L2R optimization



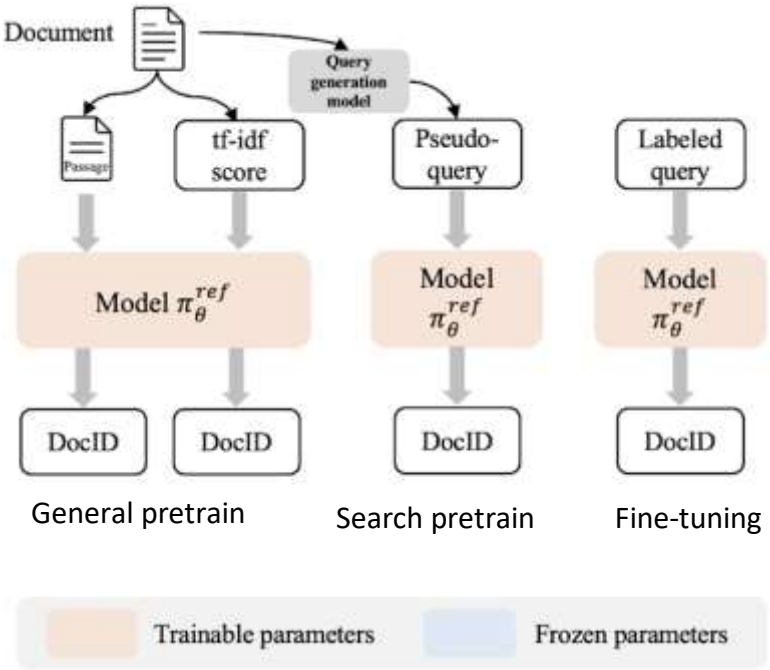
Step 1: DocID Construction

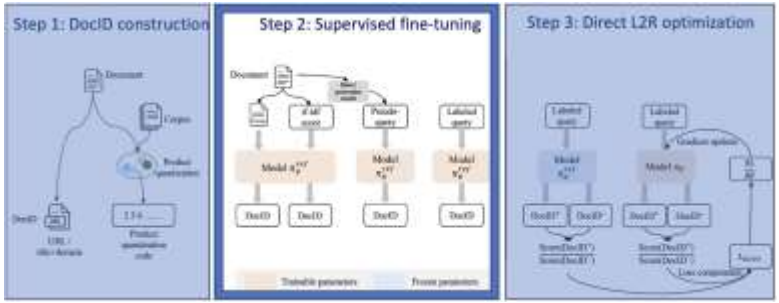




- Each document is assigned an ID via either:
  - Title-URL (lexical cues from reversed URL or title+domain)
  - PQ code (T5 embeddings quantized into semantic tokens)
- These IDs define the generator's output space during inference.

Step 2: Supervised Fine-tuning

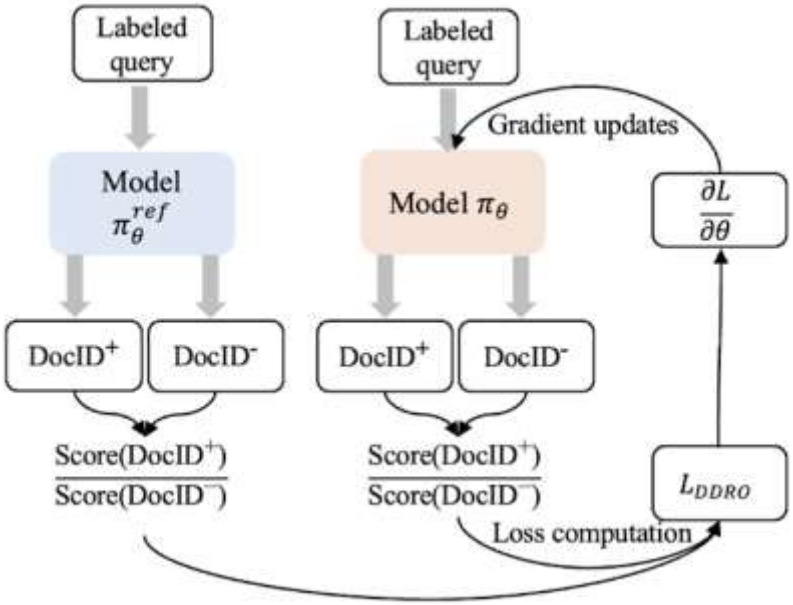


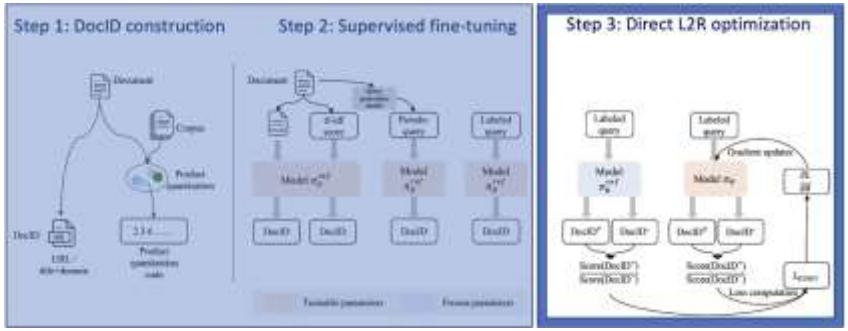


- Train T5-base model for both indexing & retrieval tasks.
- Objective: Token-level MLE with teacher forcing.

$$\max_{\pi} \mathbb{E}_{(q,\text{docid})\sim\mathcal{D}}\left[\log \pi\left(\text{docid} \mid q\right)\right]$$

Step 3: Direct L2R Optimization





- Start from frozen π^{ref}; rank documents directly using ⟨q, docID$^+$, docID$^-$⟩ triples.

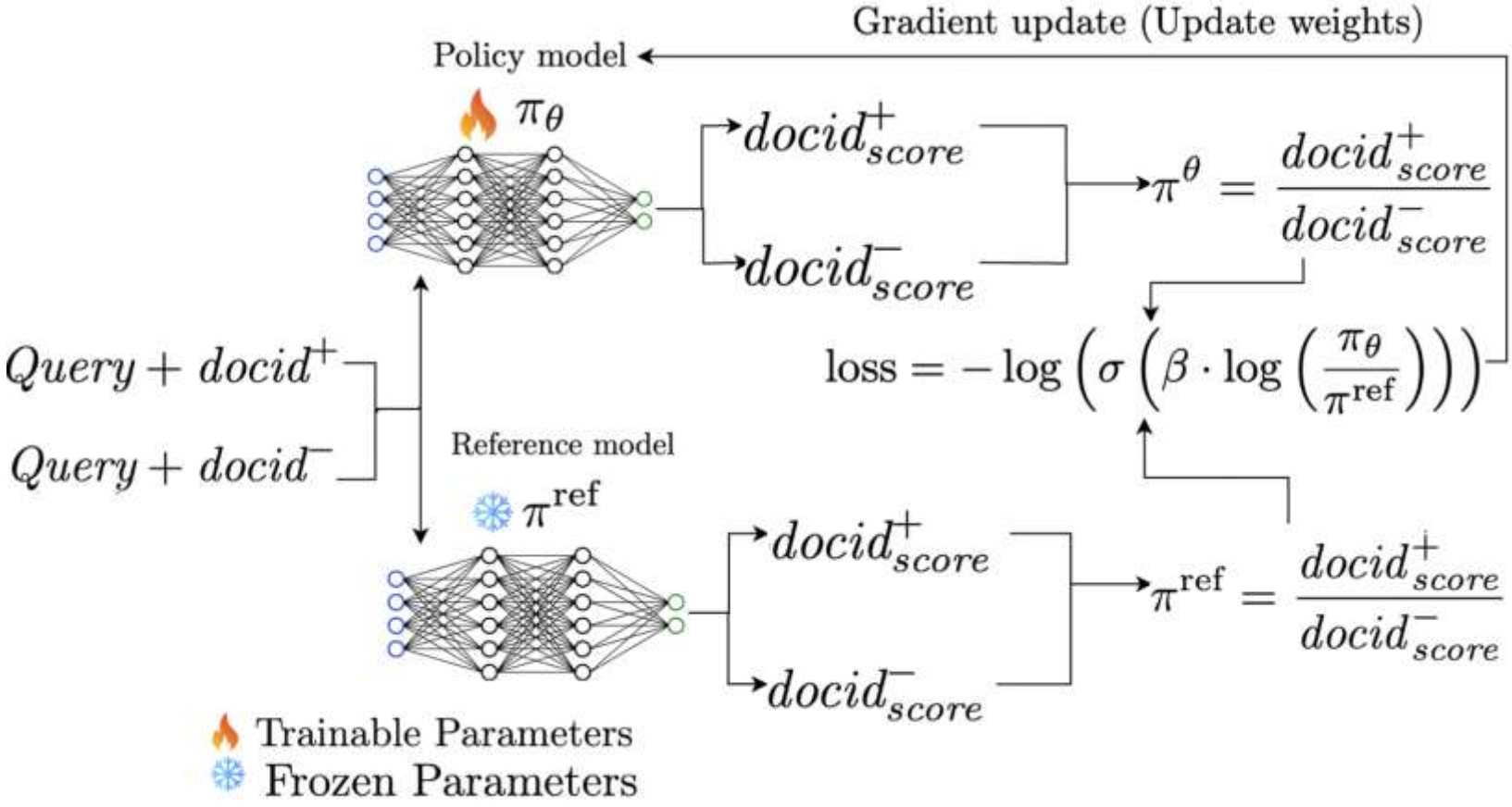
Direct L2R Optimization Using Relevance Feedback

- We adopt a KL-constrained, **DPO-style loss** (Rafailov et al., NeurIPS 2023).

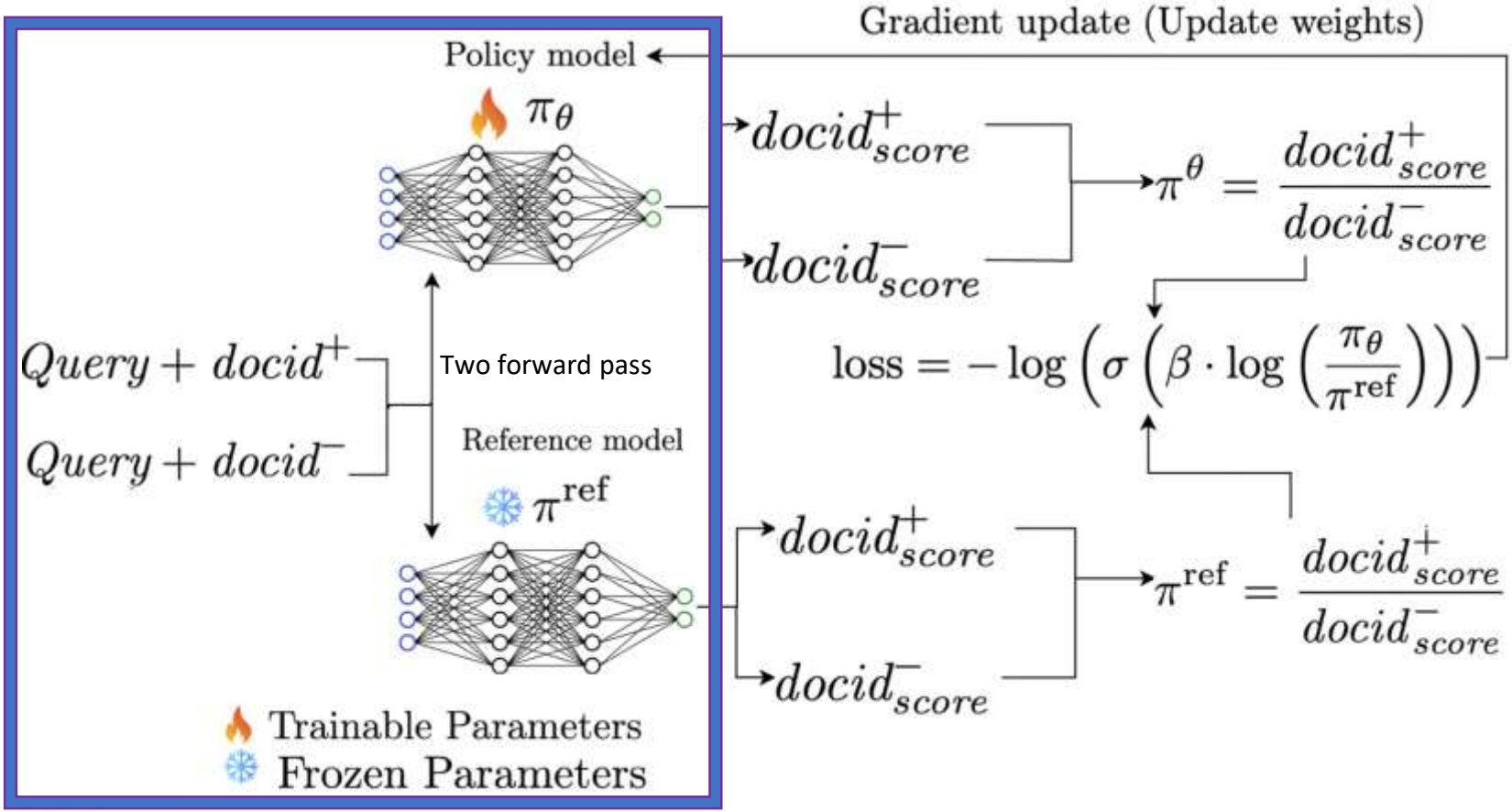
$$\mathcal{L}_{\mathrm{DDRO}}\left(\pi_\theta;\, \pi^{\mathrm{ref}}\right) = -\,\mathbb{E}_{(q,\mathrm{docid}^+,\mathrm{docid}^-)\sim\mathcal{D}} \left[\log \sigma\Big(\beta\big(\log \frac{\pi_\theta(\mathrm{docid}^+|q)}{\pi^{\mathrm{ref}}(\mathrm{docid}^+|q)} - \log \frac{\pi_\theta(\mathrm{docid}^-|q)}{\pi^{\mathrm{ref}}(\mathrm{docid}^-|q)}\big)\Big)\right].$$

- **Directly optimizes relevance ordering**, encouraging π_θ to score relevant docids higher than irrelevant ones, relative π^{ref}.

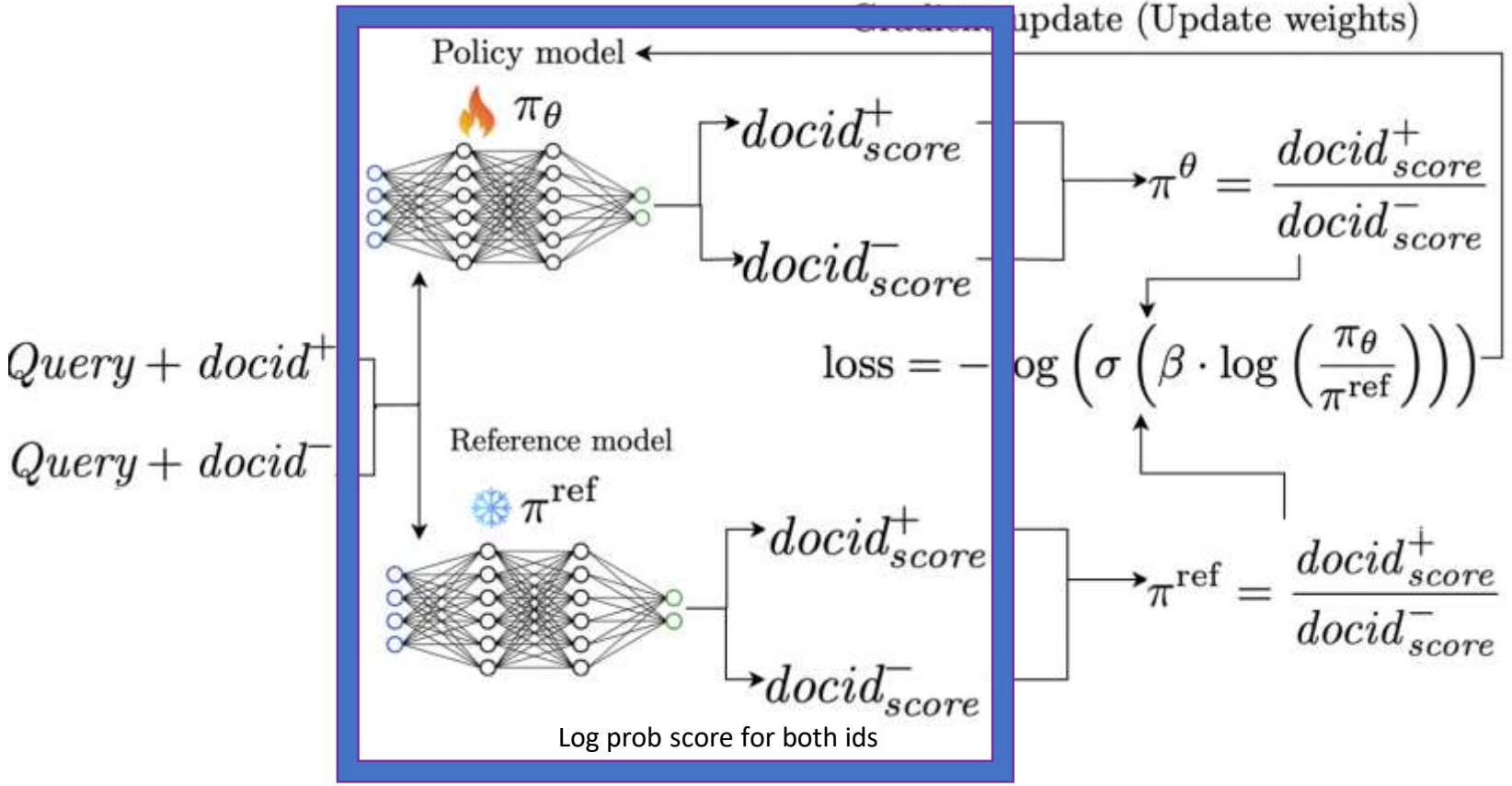
DDRO Architecture



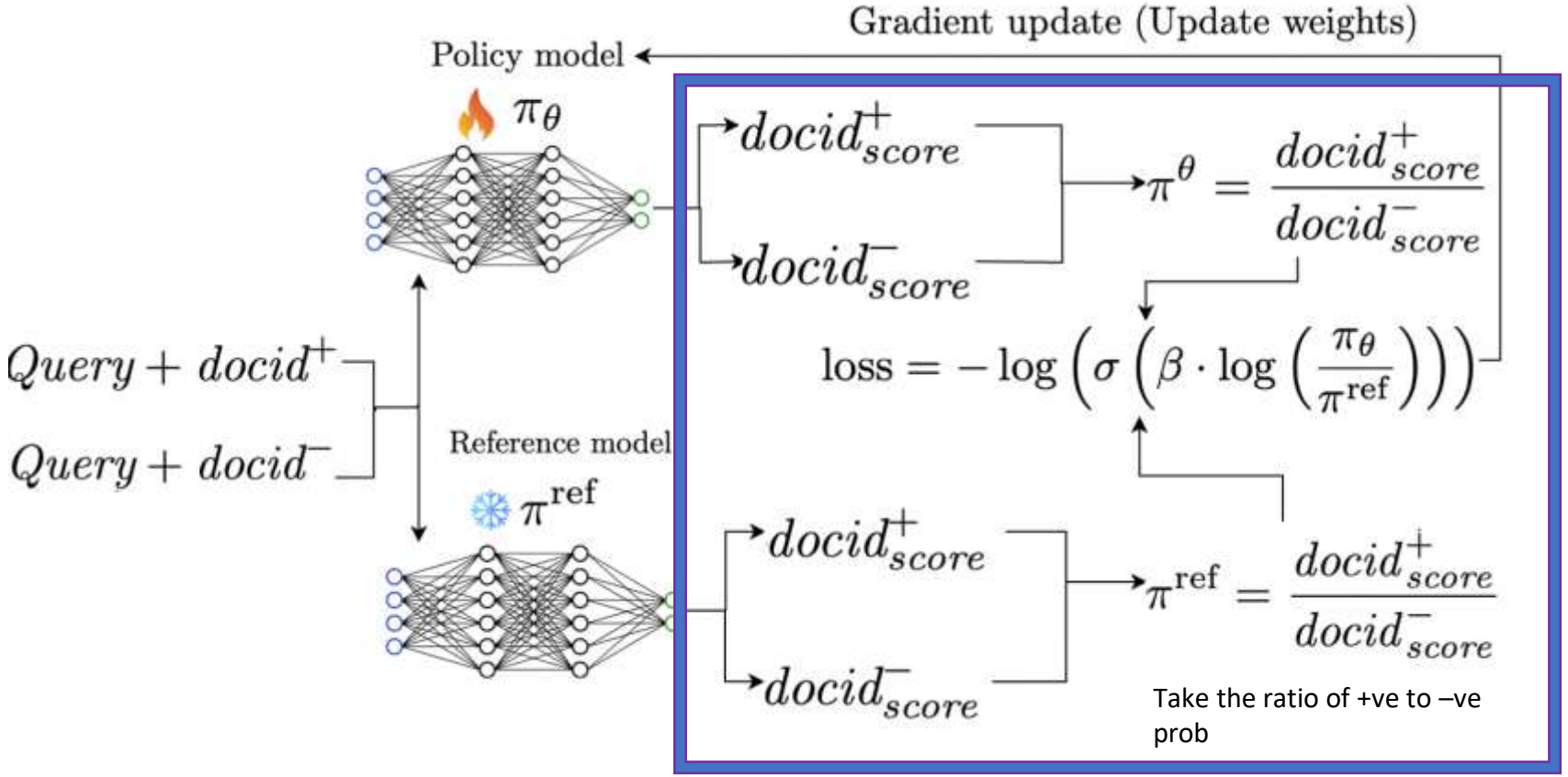
DDRO Architecture



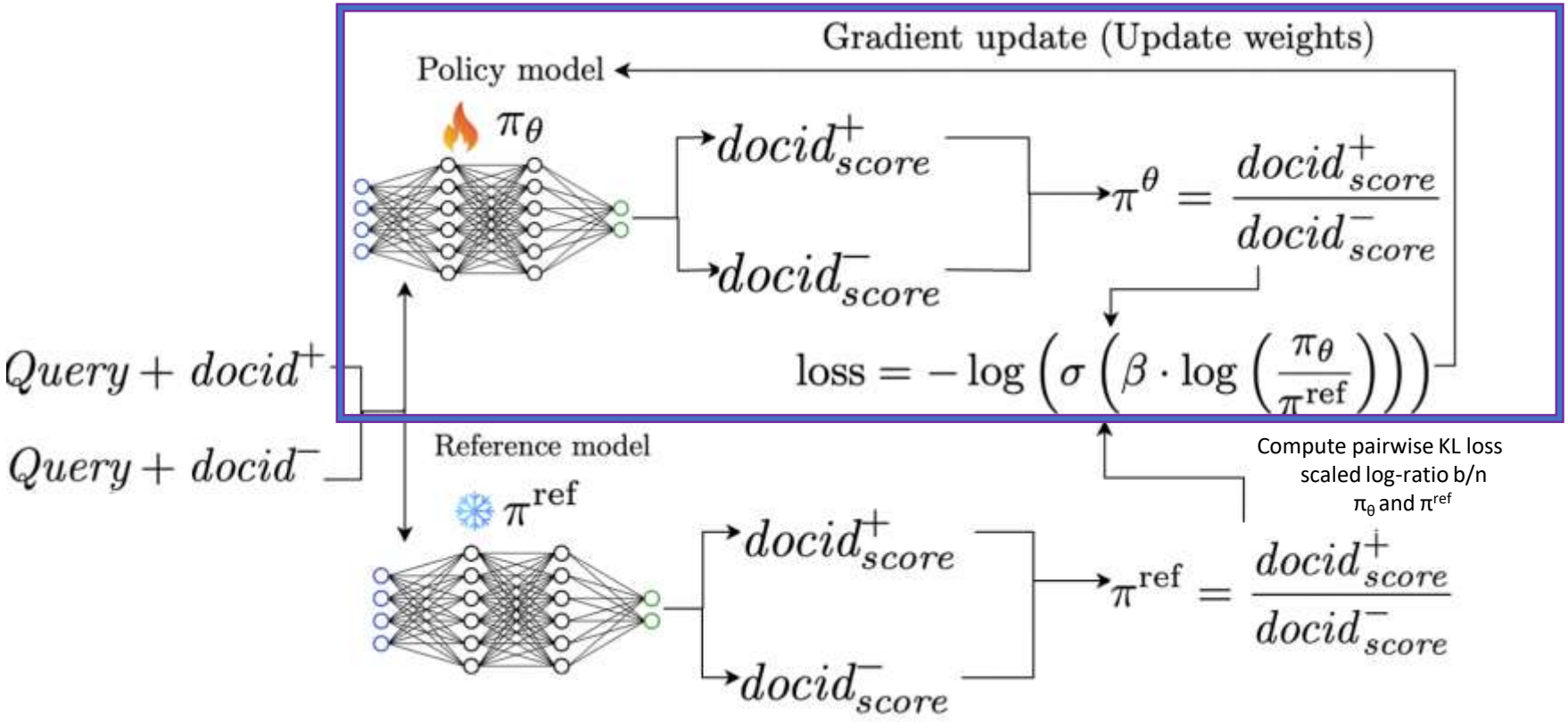
DDRO Architecture



DDRO Architecture



DDRO Architecture



Main Results

MS MARCO-doc (MS300K) ⇒ DDRO vs GenRRL

RQ1. *How does direct L2R fine-tune compare with the RLRF-based methods, such as GenRRL on the MS MARCO-doc benchmark?*

Model	R@1	R@5	R@10	MRR@10
GenRRL (TU) [76]	33.01	63.62	74.91	45.93
GenRRL (Sum) [76]	33.23	64.48	**75.80**	46.62
DDRO (PQ)	32.92	64.36	73.02	45.76
DDRO (TU)	**38.24**	**66.46**	74.01	**50.07**

- **TU advantage (hypothesis):** MS MARCO queries are short & keyword-heavy, so surface-level Title-URL IDs likely match them better than PQ codes, whose latent vectors can pick up noise from web pages.

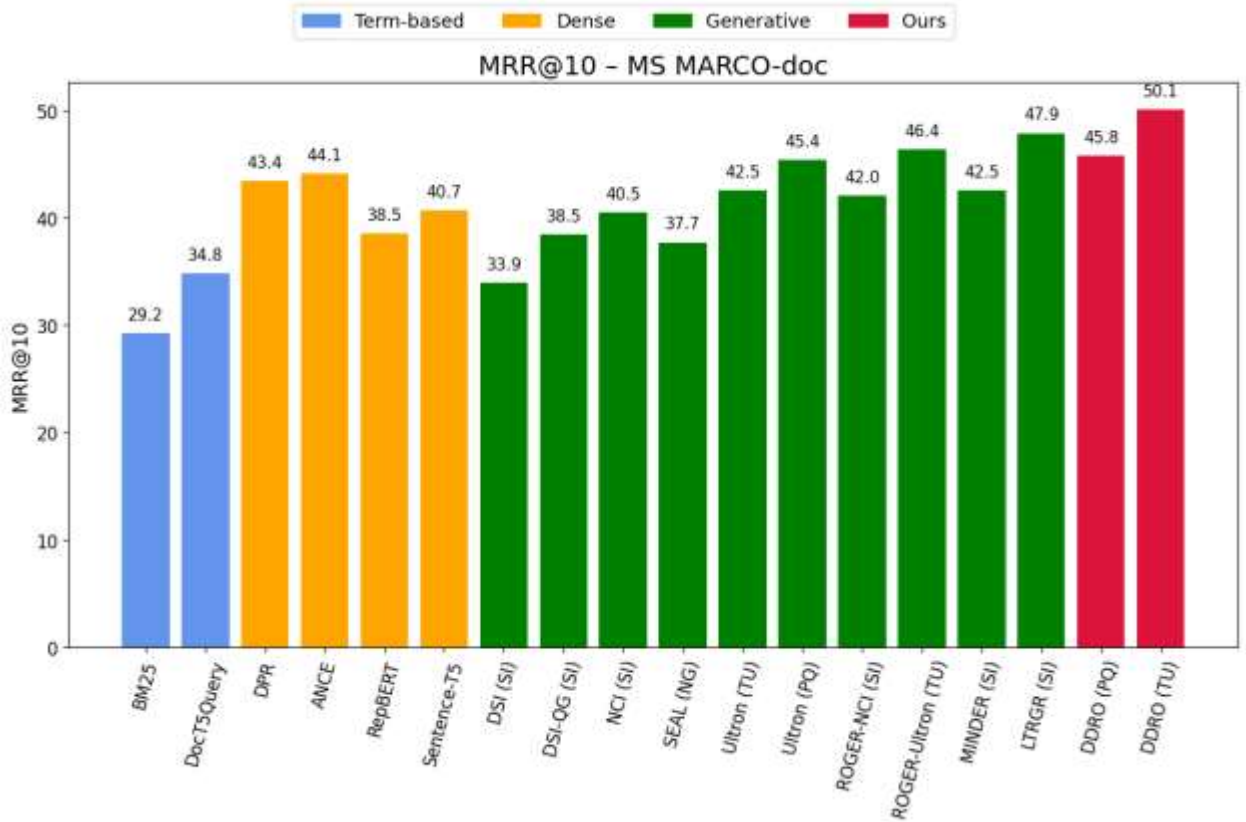
Main Results

Natural Questions 320K ⇒ DDRO vs GenRRL

Model	R@1	R@5	R@10	MRR@10
GenRRL (TU) [76]	35.79	56.49	70.96	45.73
GenRRL (Sum) [76]	36.32	57.42	**71.49**	46.31
DDRO (TU)	40.86	53.12	55.98	45.99
DDRO (PQ)	**48.92**	**64.10**	67.31	**55.51**

- **PQ advantage (hypothesis)**: The latent cluster tokens preserve Wikipedia's rich semantics, which suits the longer, knowledge-intensive NQ queries. By contrast, Title-URL IDs often collapse to a generic "wiki/…", losing discriminative detail.

RQ2. *How does DDRO perform relative to other established baselines?*



Limitations & Future Work

- Extend to larger, multilingual, and domain-specific corpora.
- Move beyond binary pairs to graded or list-wise supervision.
- Add hard-negative mining; explore multi-objective training.
- Compare with scalable Dense/GenIR systems on billion-doc testbeds.

Conclusion

Lightweight optimisation	Strong Performance	What matters
○ KL-constrained pairwise loss replaces RL + reward nets, keeping training end-to-end and simple.	○ Matches / beats heavier baselines without dense-teacher distillation or extra signals.	○ Choose IDs that fit the corpus (Title-URL for noisy web, PQ codes for Wikipedia). ○ Keep the KL pairwise loss , it drives relevance alignment.

Thank You!
Q & A