Ragnarök: A Reusable RAG Framework and Baselines for TREC 2024 Retrieval-Augmented Generation Track

Ronak Pradeep*, Nandan Thakur*, Sahel Sharifymoghaddam, Eric Zhang, Ryan Nguyen,
Daniel Campos, Nick Craswell, Jimmy Lin

University of Waterloo   Snowflake Inc.   Microsoft

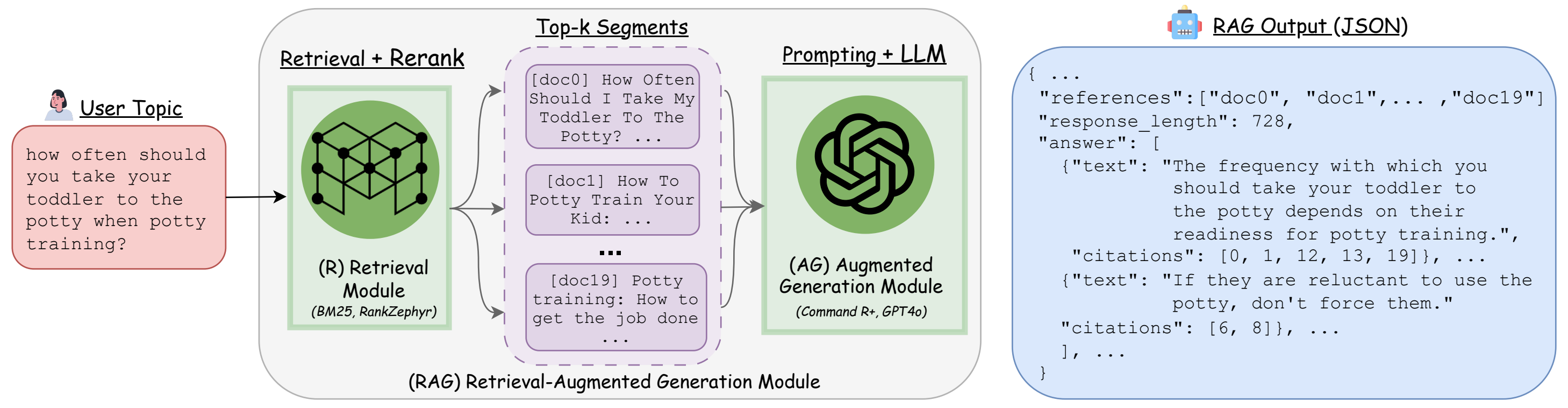
The Challenge: Standardizing RAG Evaluation

Retrieval-Augmented Generation (RAG) powers modern search (Bing, ChatGPT, Google AI Overviews), but lacks standardized frameworks for research and evaluation. The RAG community needs tools!

We introduce **Ragnarök**: an open-source RAG inference framework built with reproducibility in mind. Ragnarök has successfully been used as a SOTA pipeline in several TREC tracks — RAG, NeuCLIR, and BioGen.

Additionally, we present the MS MARCO V2.1 Collection and two topic sets to power the TREC 2024/2025 RAG Tracks.

The Ragnarök Framework


User Topic
how often should you take your toddler to the potty when potty training?
Retrieval + Rerank
(R) Retrieval Module
(BM25, RankZephyr)
Top-k Segments
[doc0] How Often Should I Take My Toddler To The Potty? ...
[doc1] How To Potty Train Your Kid: ...
...
[doc19] Potty training: How to get the job done ...
Prompting + LLM
(AG) Augmented Generation Module
(Command R+, GPT4o)
(RAG) Retrieval-Augmented Generation Module
RAG Output (JSON)

```
{ ...
 "references":["doc0", "doc1",... ,"doc19"]
 "response_length": 728,
 "answer": [
   {"text": "The frequency with which you
             should take your toddler to
             the potty depends on their
             readiness for potty training.",
    "citations": [0, 1, 12, 13, 19]}, ...
   {"text": "If they are reluctant to use the
             potty, don't force them."
   "citations": [6, 8]}, ...
   ], ...
}
```

Ragnarök provides an end-to-end RAG pipeline:

- **(R) Retrieval**: Integrates `pyserini` for first-stage retrieval & `rank_llm` for reranking (monoT5, RankZephyr, RankGPT).
- **(AG) Augmented Generation**: Supports several proprietary LLM endpoints (OpenAI, Cohere, Gemini) and open-weight LLMs (via vLLM) with configurable prompting for cited answer generation.
- **Standardized I/O**: I/O specs with sentence-level citations.
- **Tools**: REST APIs & WebUI for easy use and evaluation.

Foundation: Corpus & Topics for TREC RAG

MS MARCO V2.1 Corpora

- Deduplicated MS MARCO V2 documents (reduced by 8%).
- Segmented corpus using a sliding window (10-sentence window, 5-sentence stride) resulting in **>113 million segments**.
- Results in a retrieval chunk size suitable for both eval & RAG.

Topic Sets: Designed to challenge RAG systems beyond factoid QA.

- **TREC-RAGgy 2024**: 120 topics filtered from TREC DL '21-'23. Focus on long-form answers and information aggregation (*24% aggregation topics*). Includes mapped relevance judgments.
- **TREC-Researchy 2024**: Fresh, diverse, non-factoid topics sampled from Researchy Questions. High emphasis on knowledge intensity queries (*80%*) and multi-facetedness (*76%*). Includes relevance judgments from TREC 2024 RAG.

These resources form the basis for rigorous retrieval and RAG benchmarking in the TREC RAG Track.

Retrieval Baselines

We provide several state-of-the-art retrieval pipelines:

- **First-stage**: BM25, Dense (GTE, ArcticEmbed), Hybrid (RRF).
- **Reranking**: Pointwise (monoT5), Listwise (RankZephyr, RankGPT).

Model	nDCG@10	MAP@100	Recall@100
Lexical & Dual Encoders			
(1a) BM25	0.4227	0.1561	0.2807
(1b) Hybrid	0.6064	0.2592	0.3990
Rerankers			
(2a) RRF(1b, monoT5-3B)	0.6175	0.2708	0.4208
(2b) RRF(2a, RankZephyr)	0.6357	0.2770	0.4208

Table 1. Baseline results on Document Ranking of the TREC-RAGgy 2024 set

Augmented Generation Baselines

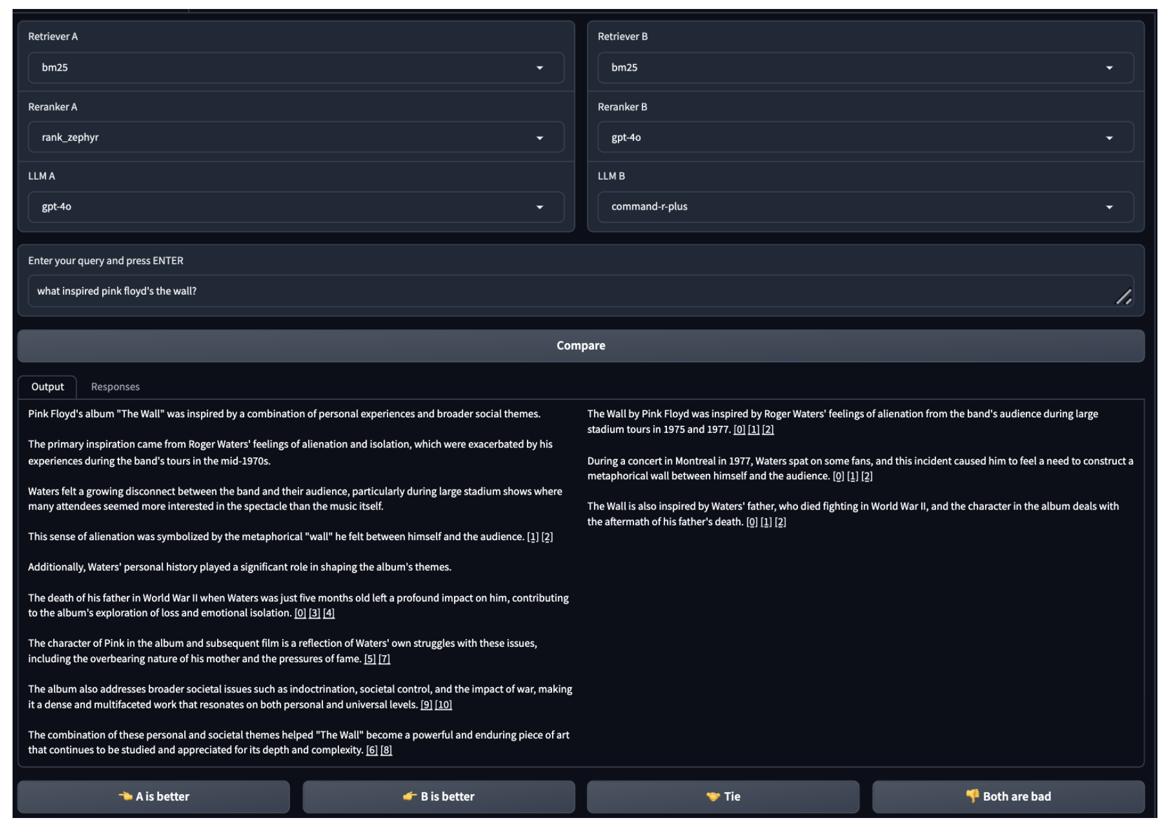
Using the retrieved context, Ragnarök generates cited answers with leading LLMs like GPT-4o, Command R+, LLaMA3.3-70B.

Configurable prompting (e.g., ChatQA-style) enables quick iterations to get consistent, well-cited answers following the output requirements of popular TREC 2024/2025 Tracks.

Qualitative analysis indicates GPT-4o is more detailed, with fewer yet more relevant citations compared to Command R+.

Ragnarök System Arena: Interactive Evaluation

Evaluate and compare RAG pipelines head-to-head!


Retriever A: bm25
Reranker A: rank_zephyr
LLM A: gpt-4o

Retriever B: bm25
Reranker B: gpt-4o
LLM B: command-r-plus

Enter your query and press ENTER: what inspired pink floyd's the wall?

Pink Floyd's album "The Wall" was inspired by a combination of personal experiences and broader social themes.

The primary inspiration came from Roger Waters' feelings of alienation and isolation, which were exacerbated by his experiences during the band's tours in the mid-1970s.

Waters felt a growing disconnect between the band and their audience, particularly during large stadium shows where many attendees seemed more interested in the spectacle than the music itself.

This sense of alienation was symbolized by the metaphorical "wall" he felt between himself and the audience. [1] [2]

Additionally, Waters' personal history played a significant role in shaping the album's themes.

The death of his father in World War II when Waters was just five months old left a profound impact on him, contributing to the album's exploration of loss and emotional isolation. [0] [3] [4]

The character of Pink in the album and subsequent film is a reflection of Waters' own struggles with these issues, including the overbearing nature of his mother and the pressures of fame. [5] [7]

The album also addresses broader societal issues such as indoctrination, societal control, and the impact of war, making it a dense and multifaceted work that resonates on both personal and universal levels. [9] [10]

The combination of these personal and societal themes helped "The Wall" become a powerful and enduring piece of art that continues to be studied and appreciated for its depth and complexity. [6] [8]

The Wall by Pink Floyd was inspired by Roger Waters' feelings of alienation from the band's audience during large stadium tours in 1975 and 1977. [0] [1] [2]

During a concert in Montreal in 1977, Waters spat on some fans, and this incident caused him to feel a need to construct a metaphorical wall between himself and the audience. [0] [1] [2]

The Wall is also inspired by Waters' father, who died fighting in World War II, and the character in the album deals with the aftermath of his father's death. [0] [1] [2]

- **Pairwise Comparison**: Judge outputs from two pipelines side-by-side (blinded or unblinded).
- **ELO Leaderboards**: Rank individual modules (Retrieval, Generation) or full RAG pipelines.

Check it Out!

ECIR 2025, Lucca, Italy

{ronak.pradeep, nandan.thakur, jimmylin}@uwaterloo.ca