SOVEREIGN: Promptriever: Instruction-Trained Retrievers Can Be Prompted Like Language Model

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

Instruction-tuned language models (LM) are able to respond to imperative commands, providing a more natural user interface compared to their base counterparts. In this work, we present Promptriever, the first retrieval model able to be prompted like an LM. To train Promptriever, we curate and release a new instance-level instruction training set from MS MARCO, spanning nearly 500k instances. Promptriever not only achieves strong performance on standard retrieval tasks, but also follows instructions. We observe: (1) large gains (reaching SoTA) on following detailed relevance instructions (+14.3

1 Introduction

Analysis of: Promptriever: Instruction-Trained Retrievers Can Be Prompted Like Language Models. Research goal: Can adversarial training on multi-hop queries from MuSiQue improve retriever robustness on out-of-domain BEIR subsets (e.g., SciFact, TREC-COVID) compared to single-context adversarial training, measured by MRR@10?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

3 papers retrieved. 5 claims extracted, 5 verified. Tribunal: 7.5/10 → APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
Promptriever is the first retrieval model able to be prompted like a language model.	✓	0.27
Promptriever achieves large gains (reaching SoTA) on following detailed relevance instructions (+14.3 p-MRR / +3.1 nDCG	✓	0.30
Promptriever shows significantly increased robustness to lexical choices/phrasing in the query+instruction (+12.9 Robust	✓	0.28
Promptriever can perform hyperparameter search via prompting to reliably improve retrieval performance (+1.4 average inc	✓	0.30
The instruction training set for Promptriever was curated from MS MARCO and spans nearly 500k instances.	✓	0.18

References

- https://www.semanticscholar.org/paper/7b3539219818f3be0b6b52a6131edbfe2346bf45
- https://www.semanticscholar.org/paper/27a3ef52f4549ad202ea88e9cadefccf73cf4d8c
- https://www.semanticscholar.org/paper/dd32434642801a9702c51f6465615fb05e191d26