

SOVEREIGN: DRAGON: Domain-specific Robust Automatic Data Generation for RAG Optimization

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

Retrieval-augmented generation (RAG) can substantially enhance the performance of LLMs on knowledge-intensive tasks. Various RAG paradigms - including vanilla, planning-based, and iterative RAG - all depend on a robust retriever, yet existing retrievers rely heavily on public knowledge and often falter when faced with domain-specific queries. To address these limitations, we introduce DRAGON, a framework that combines a data-construction modeling approach with a scalable synthetic data-generation pipeline, specifically designed to optimize domain-specific retrieval performance and bolster retr

1 Introduction

Analysis of: DRAGON: Domain-specific Robust Automatic Data Generation for RAG Optimization. Research goal: How does the number of hops in multi-hop queries (e.g., 2-hop vs 3-hop from HotPotQA) affect the MRR@10 improvement from adversarial training on retriever robustness in RAG systems?.

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

18 papers retrieved. 4 claims extracted, 4 verified. Tribunal: 8.2/10 → 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
DRAGON is a framework that combines a data-construction modeling approach with a scalable synthetic data-generation pipe	✓	0.42
DRAGONBench is a benchmark spanning 8 domain-specific document collections across 4 distinct fields with a wide spectrum	✓	0.35
Retrievers trained on data generated by DRAGON yield significant performance gains and exhibit strong cross-domain gener	✓	0.29
When DRAGON-optimized retrievers are integrated into vanilla, planning-based, and iterative RAG paradigms, consistent en	✓	0.36

References

- <https://www.semanticscholar.org/paper/f005b861225d491a17f4ea7385b8592cdea01cbe>
- <https://www.semanticscholar.org/paper/8efc7081a7d63ae3161e9e1651f625eea371ee1f>
- <https://www.semanticscholar.org/paper/5a6e848e7c8226ebbeecddd44445d5f9f0230fe0>