

SOVEREIGN: An Exploration of Data Augmentation and Sampling Techniques for Domain-Agnostic

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 27, 2026

Abstract

To produce a domain-agnostic question answering model for the Machine Reading Question Answering (MRQA) 2019 Shared Task, we investigate the relative benefits of large pre-trained language models, various data sampling strategies, as well as query and context paraphrases generated by back-translation. We find a simple negative sampling technique to be particularly effective, even though it is typically used for datasets that include unanswerable questions, such as SQuAD 2.0. When applied in conjunction with per-domain sampling, our XLNet (Yang et al., 2019)-based submission achieved the second

1 Introduction

Analysis of: An Exploration of Data Augmentation and Sampling Techniques for Domain-Agnostic Question Answering. Research goal: How does negative sampling scale with model size (e.g., 7B vs 70B) in domain-agnostic QA when evaluated on out-of-distribution benchmarks like MRQA 2019, and can it match the performance of domain-specific fine-tuning?.

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

11 papers retrieved. 2 claims extracted, 2 verified. Tribunal: 7.2/10 → 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
A simple negative sampling technique is particularly effective for domain-agnostic question answering, even though it is	✓	0.47
When applied in conjunction with per-domain sampling, an XLNet-based submission achieved the second best Exact Match and	✓	0.44

References

- <http://arxiv.org/abs/1910.09342v2>
- <http://arxiv.org/abs/2404.14700v4>
- <http://arxiv.org/abs/1912.02145v1>