SOVEREIGN: How do different embedding models (SPECTER, ConRetri(Saltz)) influence RAG performance on the Natural Question

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

Despite progress in perceptual tasks such as image classification, computers still perform poorly on cognitive tasks such as image description and question answering. Cognition is core to tasks that involve not just recognizing, but reasoning about our visual world. However, models used to tackle the rich content in images for cognitive tasks are still being trained using the same datasets designed for perceptual tasks. To achieve success at cognitive tasks, models need to understand the interactions and relationships between objects in an image. When asked "What vehicle is the person riding?"

1 Introduction

Analysis of: Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. Research goal: How do different embedding models (SPECTER, ConRetri(Saltz)) influence RAG performance on the Natural Questions benchmark?.

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

2 papers retrieved. 7 claims extracted, 7 verified. Tribunal: 8.7/10 $\rightarrow$ 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
The Visual Genome dataset contains over 108K images	✓	0.21
Each image in Visual Genome has an average of 35 objects	✓	0.17
Each image in Visual Genome has an average of 26 attributes	✓	0.16
Each image in Visual Genome has an average of 21 pairwise relationships between objects	✓	0.21
Computers perform poorly on cognitive tasks such as image description and question answering	✓	0.28
Models for cognitive tasks need to understand interactions and relationships between objects in an image	✓	0.29
The example question 'What vehicle is the person riding?' requires identifying relationships riding(man, carriage) and p	✓	0.27

References

- https://doi.org/10.18653/v1/d19-1410
- https://doi.org/10.1007/s11263-016-0981-7