Scaling Pre-Training Data and Distributional Robustness in Dense Retrieval Models

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the scaling of pre-training data volume affect the distributional robustness of dense retrieval models when evaluated on cross-domain reasoning benchmarks versus in-domain fine-tuned. 7 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Pre-training vs. Fine-tuning: A Reproducibility Study on Dense Retrieval Knowledge Acquisition. Research question: How does the scaling of pre-training data volume affect the distributional robustness of dense retrieval models when evaluated on cross-domain reasoning benchmarks versus in-domain fine-tuned baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

13 papers retrieved. 7 claims extracted; 1 independently verified. Quality review score: 4.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The study investigates whether findings from DPR fine-tuning with CLS token representations apply to the Contriever fine	✓	0.18
The study investigates whether findings obtained for the BERT backbone transfer to decoder-based backbones.	×	0.09
The study conducts a linear probing experiment following the methodology proposed by Reichman and Heck [2024].	×	0.02
Linear probing involves training a simple linear classifier on embeddings extracted from different layers of a model.	×	0.03
The task used in the Reichman and Heck [2024] linear probing experiment is distinguishing relevant and irrelevant docume	×	0.02
The dense retrieval task involves training related to establishing the similarity between a query and a passage, which d	×	0.06
High accuracy achieved by a linear classifier in a probing experiment suggests that the embeddings contain discriminativ	×	0.02

References

- http://arxiv.org/abs/2410.21676v4
- http://arxiv.org/abs/2603.02208v1
- http://arxiv.org/abs/2505.07166v1