Manifold-Aware Dense Retrieval Efficiency vs. DPR on Large-Scale Passage Retrieval

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the inference efficiency of manifold-aware dense retrieval models compare to baseline DPR models on large-scale passage retrieval tasks (e.g., MS MARCO) when using approximate nearest. The deployment of large language models (LLMs) within the healthcare sector has sparked both enthusiasm and apprehension. These models exhibit the remarkable ability to provide proficient responses to free-text queries, demonstrating a nuanced understanding of professional. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Large Language Models in Healthcare and Medical Domain: A Review. Research question: How does the inference efficiency of manifold-aware dense retrieval models compare to baseline DPR models on large-scale passage retrieval tasks (e.g., MS MARCO) when using approximate nearest neighbor search with manifold-aware distance metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

15 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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.

References

- https://doi.org/10.1007/s11831-016-9206-z
- https://doi.org/10.3390/informatics11030057
- https://doi.org/10.18653/v1/2022.findings-acl.316