Multimodal vs. Text-Only Dense Retrieval Models Under Spelling Errors

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How do recent multimodal dense retrieval models (e.g., M3-Rec) perform in misspelling robustness tasks compared to text-only models when evaluated on benchmarks like Flickr30k or XLENT with induced. Dialogue systems powered by large language models (LLMs) show strong generative abilities but often struggle with informal language, long-term coherence, and grounded responses in expert-driven conversations. This thesis presents three complementary methods to address these. 9 claims were extracted from source literature; 9 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Zamansal balamsal konu tahmini ve kontroll yant getirme ile diyalog sistemlerinin gelitirilmesi. Research question: How do recent multimodal dense retrieval models (e.g., M3-Rec) perform in misspelling robustness tasks compared to text-only models when evaluated on benchmarks like Flickr30k or XLENT with induced spelling errors?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.8/10.

3 Results

4 papers retrieved. 9 claims extracted; 9 independently verified. Quality review score: 8.8/10.

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
Dialogue systems powered by large language models (LLMs) show strong generative abilities but often struggle with inform	✓	0.39
A novel temporal forecasting framework is introduced that models dialogue topic trajectories as time series and predicts	✓	0.37
The temporal forecasting framework offers interpretable and accurate forecasting in both domain-specific and open-domain	✓	0.23
SALDIRAY is a task-agnostic standardization pipeline that normalizes messages affected by spelling errors, slang, and ab	✓	0.27
SALDIRAY improves downstream NLP performance across tasks such as sentiment analysis and topic labeling.	✓	0.21
RAP (Retrieval-Augmented Paraphrasing) is a retrieval-based generation method that retrieves similar past responses and	✓	0.35
RAP significantly reduces hallucinations while maintaining stylistic alignment and relevance.	✓	0.15
The methods presented improve the robustness, foresight, and factual grounding of conversational agents.	✓	0.19
The methods advance the reliability of LLM-based dialogue systems in real-world applications.	✓	0.22

References

- https://doi.org/10.36227/techrxiv.176282213.31303325/v1
- https://doi.org/10.48550/arxiv.2404.14890
- https://openalex.org/W7139877339