Multimodal vs. Text-Only Models in Dialectal Cultural Comprehension on AraDiCE

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How do multimodal models (e.g., vision-language models) perform on dialectal cultural comprehension tasks compared to text-only models, evaluated by accuracy on the AraDiCE benchmark when. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: AraDiCE: Benchmarks for Dialectal and Cultural Capabilities in LLMs. Research question: How do multimodal models (e.g., vision-language models) perform on dialectal cultural comprehension tasks compared to text-only models, evaluated by accuracy on the AraDiCE benchmark when incorporating dialect-specific visual cues (e.g., regional symbols, cultural artifacts)?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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.

References

- http://arxiv.org/abs/2404.07214v4
- http://arxiv.org/abs/2409.11404v3
- http://arxiv.org/abs/2510.06371v3