

Synthetic Data Size and Post-Editing Quality in LLM Dialect Comprehension on AraDiCE

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the impact of synthetic dataset size and human post-editing quality on the dialect comprehension capabilities of LLMs, as measured by accuracy on the AraDiCE benchmark when comparing models. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. 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: What is the impact of synthetic dataset size and human post-editing quality on the dialect comprehension capabilities of LLMs, as measured by accuracy on the AraDiCE benchmark when comparing models trained on varying proportions of machine-translated vs. human-edited data?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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/2508.13525v1>
- <http://arxiv.org/abs/2409.11404v3>
- <http://arxiv.org/abs/2512.21787v2>