

Publicly Available Datasets and Performance Gaps in LLaMA on Rare Language Tasks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: Does the exclusive use of publicly available datasets in LLaMA training result in measurable performance gaps on rare language tasks in XTREME compared to models trained on proprietary corpora. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Evaluating the Limits of Large Language Models in Multilingual Legal Reasoning. Research question: Does the exclusive use of publicly available datasets in LLaMA training result in measurable performance gaps on rare language tasks in XTREME compared to models trained on proprietary corpora?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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/2509.22472v1>
- <http://arxiv.org/abs/2409.13897v1>
- <http://arxiv.org/abs/2506.15415v1>