

Targeted Lexical Injection in Lughha-Llama for Zero-Shot XNLI Generalization Across Low-Resource Languages

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: Does the latent cross-lingual alignment achieved by Targeted Lexical Injection in Lughha-Llama generalize to zero-shot performance on the XNLI dataset for unrelated low-resource language families. 9 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Targeted Lexical Injection: Unlocking Latent Cross-Lingual Alignment in Lughha-Llama via Early-Layer LoRA Fine-Tuning. Research question: Does the latent cross-lingual alignment achieved by Targeted Lexical Injection in Lughha-Llama generalize to zero-shot performance on the XNLI dataset for unrelated low-resource language families?.

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 4.2/10.

3 Results

10 papers retrieved. 9 claims extracted; 2 independently verified. Quality review score: 4.2/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
Layer 0 (input embeddings) showed a modest average cosine similarity of approximately 0.3153.	×	0.08
Layer 1 showed an average cosine similarity of 0.9808.	×	0.09
Layer 2 exhibited the peak average cosine similarity, reaching 0.99998.	×	0.09
Layer 31 showed an average similarity of 0.9876 in the pilot scan.	×	0.04
The baseline output similarity observed on the full evaluation set was approximately 0.32.	×	0.10
The average cosine similarity at the final output layer (Layer 31) of the base model was approximately 0.3211 for the tr	✓	0.17
The base model used is Lughu-Llama-8B-wura, an open-source LLM adapted for several African languages, including Swahili,	×	0.11
The model is loaded in 4-bit precision using bitsandbytes with NF4 quantization and torch.bfloat16 as the compute data t	×	0.02
The pilot study identified Layer 2 as exhibiting the highest degree of inherent cross-lingual lexical alignment for Swah	✓	0.27

References

- <http://arxiv.org/abs/2309.10891v1>
- <http://arxiv.org/abs/2506.15415v1>
- <http://arxiv.org/abs/2310.09917v3>