

How does the scaling of model size (e.g., 7B vs. 13B vs. 30B parameters) affect the cross-lingual transferability

Assignee Research

June 10, 2026

Abstract

Multilingual large-scale Pretrained Language Models (PLMs) have been shown to store considerable amounts of factual knowledge, but large variations are observed across languages. With the ultimate goal of ensuring that users with different language backgrounds obtain consistent feedback from the same model, we study the cross-lingual consistency (CLC) of factual knowledge in various multilingual PLMs. To this end, we propose a Ranking-based Consistency (RankC) metric to evaluate knowledge consistency across languages independently from accuracy. Using this metric, we conduct an in-depth analysis

1 Introduction

This paper examines: Cross-Lingual Consistency of Factual Knowledge in Multilingual Language Models. Research question: How does the scaling of model size (e.g., 7B vs. 13B vs. 30B parameters) affect the cross-lingual transferability of prosody detection in multimodal LLMs, as evaluated on non-English subsets of MMSU and other benchmarks like VoxCeleb or CREMA-D?.

2 Methodology

Systematic literature search across multiple databases yielded 13 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

13 papers retrieved. 16 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.

5 Extracted Claims

Claim	Verified	Confidence
The multilingual BLOOM-3b model outputs consistently correct completions for the query regarding the capital of China wh	×	0.02
The multilingual BLOOM-3b model outputs inconsistent or incorrect completions for the query regarding the capital of Chi	×	0.02
When queried about the capital of China in Hungarian, the BLOOM-3b model outputs 'Carlisle'.	×	0.01
When queried about the capital of China in Greek, the BLOOM-3b model outputs 'Berne' ($\rho\eta\nu\epsilon$).	×	0.01
The multilingual BLOOM-3b model outputs consistently incorrect completions for the query regarding the original language	×	0.04
When queried about the original language of The Godfather in Hungarian, the BLOOM-3b model outputs 'Russian' (orosz).	×	0.02
When queried about the original language of The Godfather in Greek, the BLOOM-3b model outputs 'English' ($\gamma\gamma\lambda\iota\kappa$).	×	0.03
The average Cross-Lingual Consistency (CLC) accuracy for the BLOOM-3b model across the tested languages is 25%.	×	0.14
The average Cross-Lingual Consistency (CLC) accuracy for the BLOOM-1.7b model across the tested languages is 25%.	×	0.13
The average Cross-Lingual Consistency (CLC) accuracy for the BLOOM-1.1b model across the tested languages is 24%.	×	0.13
The average Cross-Lingual Consistency (CLC) accuracy for the BLOOM-560m model across the tested languages is 23%.	×	0.12
The mT5-large model achieves a CLC accuracy of 46.8 in Russian (ru).	×	0.04
The LLaMA-7b model achieves a CLC accuracy of 56.4 in Russian (ru).	×	0.10
The BLOOM-7.1b model achieves a CLC accuracy of 51.7 in Russian (ru).	×	0.03
The BLOOM-3b model achieves a CLC accuracy of 52.3 in Russian (ru).	×	0.03
The code and data for the study are released at https://github.com/Betswish/Cross-Lingual-Consistency .	×	0.11

References

- <http://arxiv.org/abs/2310.00905v2>
- <http://arxiv.org/abs/2310.10378v5>
- <http://arxiv.org/abs/2506.04779v3>