

# How does the preservation of downstream classification accuracy compare between clustering and projection debiasing when

Assignee Research

June 11, 2026

## Abstract

Abstract Based on the foundation of Large Language Models (LLMs), Multilingual LLMs (MLLMs) have been developed to address the challenges faced in multilingual natural language processing, hoping to achieve knowledge transfer from high-resource languages to low-resource languages. However, significant limitations and challenges still exist, such as language imbalance, multilingual alignment, and inherent bias. In this paper, we aim to provide a comprehensive analysis of MLLMs, delving deeply into discussions surrounding these critical issues. First of all, we start by presenting an overview of

## 1 Introduction

This paper examines: A survey on multilingual large language models: corpora, alignment, and bias. Research question: How does the preservation of downstream classification accuracy compare between clustering and projection debiasing when evaluated on multiclass religious bias datasets?.

## 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 8.5/10.

## 3 Results

10 papers retrieved. 8 claims extracted; 8 independently verified. Quality review score: 8.5/10.

## 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
Multilingual LLMs (MLLMs) have been developed to address challenges in multilingual natural language processing, aiming	✓	0.38
Significant limitations and challenges still exist in MLLMs, such as language imbalance, multilingual alignment, and inh	✓	0.28
The paper provides a comprehensive analysis of MLLMs, discussing critical issues like language imbalance, multilingual a	✓	0.22
The paper presents an overview of MLLMs, covering their evolutions, key techniques, and multilingual capacities.	✓	0.24
The paper explores multilingual training corpora of MLLMs and multilingual datasets oriented for downstream tasks to enh	✓	0.31
The paper surveys state-of-the-art studies of multilingual representations and investigates whether current MLLMs can le	✓	0.25
The paper discusses bias on MLLMs, including its categories, evaluation metrics, and debiasing techniques.	✓	0.22
The paper discusses existing challenges and points out promising research directions of MLLMs.	✓	0.15

## References

- <https://doi.org/10.1007/s11704-024-40579-4>
- <https://doi.org/10.48550/arxiv.2404.00929>
- <https://doi.org/10.18653/v1/2022.acl-long.541>