

# Phi-3-Mini and Mistral-7B Performance Variance on Multilingual GSM-Symbolic Benchmarks

Assignee Research

June 6, 2026

## Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the performance variance of Phi-3-mini versus Mistral-7B-v0.1 on GSM-Symbolic generated instances across non-English languages compared to the original MGSM dataset. 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: Comparative Code Structure Analysis using Deep Learning for Performance Prediction. Research question: What is the performance variance of Phi-3-mini versus Mistral-7B-v0.1 on GSM-Symbolic generated instances across non-English languages compared to the original MGSM dataset?.

## 2 Methodology

Systematic literature search across multiple databases yielded 11 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

11 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/2405.19595v1>
- <http://arxiv.org/abs/2310.06825v1>
- <http://arxiv.org/abs/2102.07660v2>