

# Open-Source vs. Proprietary Multimodal Models on MME Perception Benchmarks

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: How does the perception accuracy of open-source multimodal models like LLaVA compare to proprietary models on the MME benchmark's existence and count subtasks. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: MME: A Comprehensive Evaluation Benchmark for Multimodal Large Language Models. Research question: How does the perception accuracy of open-source multimodal models like LLaVA compare to proprietary models on the MME benchmark's existence and count subtasks?.

## 2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.5/10.

## 3 Results

7 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.5/10.

## 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.

## References

- <http://arxiv.org/abs/2306.13394v5>
- <http://arxiv.org/abs/2410.02712v2>
- <http://arxiv.org/abs/2504.03641v2>