

How do multimodal models like OpenPangu-7B-MLA compare to unimodal baselines in terms of reasoning accuracy on

Assignee Research

June 10, 2026

Abstract

We propose LogicVista, an evaluation benchmark that assesses the integrated logical reasoning capabilities of multimodal large language models (MLLMs) in Visual contexts. Recent advancements in MLLMs have demonstrated various fascinating abilities, from crafting poetry based on an image to performing mathematical reasoning. However, there is still a lack of systematic evaluation of MLLMs' proficiency in logical reasoning tasks, which are essential for activities like navigation and puzzle-solving. Thus we evaluate general logical cognition abilities across 5 logical reasoning tasks encompassin

1 Introduction

This paper examines: LogicVista: Multimodal LLM Logical Reasoning Benchmark in Visual Contexts. Research question: How do multimodal models like OpenPangu-7B-MLA compare to unimodal baselines in terms of reasoning accuracy on the EchoMind benchmark when evaluated under different noise conditions using the F1 score?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.8/10.

3 Results

4 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.8/10.

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.

References

- <http://arxiv.org/abs/2604.08140v1>
- <http://arxiv.org/abs/2510.22758v2>
- <http://arxiv.org/abs/2407.04973v1>