

Large Multimodal Models Under Visual Reasoning Perturbations on HumanEval-V

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the impact of visual reasoning perturbations on the robustness of Large Multimodal Models when evaluated on the HumanEval-V benchmark. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: On the Reliability of Vision-Language Models Under Adversarial Frequency-Domain Perturbations. Research question: What is the impact of visual reasoning perturbations on the robustness of Large Multimodal Models when evaluated on the HumanEval-V benchmark?.

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 3.7/10.

3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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/2410.12381v3>
- <http://arxiv.org/abs/2507.22398v3>
- <http://arxiv.org/abs/2407.04973v1>