

VELMA Performance on Obstructed-R2R Against Flamingo and PaLI Models

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the performance of VELMA compare to other multimodal LLMs (e.g., Flamingo, PaLI) on the Obstructed-R2R benchmark in terms of success rate and path length efficiency. Large Vision-Language Models (LVLMs) have recently played a dominant role in multimodal vision-language learning. Despite the great success, it lacks a holistic evaluation of their efficacy. 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: LVLM-eHub: A Comprehensive Evaluation Benchmark for Large Vision-Language Models. Research question: How does the performance of VELMA compare to other multimodal LLMs (e.g., Flamingo, PaLI) on the Obstructed-R2R benchmark in terms of success rate and path length efficiency?.

2 Methodology

Systematic literature search across multiple databases yielded 15 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

15 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/2306.09265v1>
- <http://arxiv.org/abs/2407.17856v4>
- <http://arxiv.org/abs/2411.18711v4>