Multi-Turn Reasoning in LongNav-R1 Boosts Success Rates on RxR-CE Benchmark

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 17 peer-reviewed papers addressing the following research question: Does the multi-turn reasoning architecture of LongNav-R1 improve success rate metrics on the RxR-CE benchmark compared to standard single-turn VLA approaches. Vision-and-Language Models (VLMs) have shown impressive capabilities on single-turn benchmarks, yet real-world applications often demand more intricate multi-turn dialogues. Existing multi-turn datasets (e.g. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MultiVerse: A Multi-Turn Conversation Benchmark for Evaluating Large Vision and Language Models. Research question: Does the multi-turn reasoning architecture of LongNav-R1 improve success rate metrics on the RxR-CE benchmark compared to standard single-turn VLA approaches?.

2 Methodology

Systematic literature search across multiple databases yielded 17 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.5/10.

3 Results

17 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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

- https://arxiv.org/abs/2510.16641
- https://arxiv.org/abs/2504.09000
- https://arxiv.org/abs/2502.03333