LongNav-R1 Efficiency Gains Over Single-Turn VLA Policies on RxR-CE

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What is the efficiency gain of LongNav-R1 compared to single-turn VLA policies in terms of inference time and compute resources on the RxR-CE benchmark. Embodied AI is widely recognized as a cornerstone of artificial general intelligence (AGI) because it involves controlling embodied agents to perform tasks in the physical world. Building on the success of large language models (LLMs) and vision-language models (VLMs), a new. 9 claims were extracted from source literature; 9 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 9.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Survey on Vision–Language–Action Models for Embodied AI. Research question: What is the efficiency gain of LongNav-R1 compared to single-turn VLA policies in terms of inference time and compute resources on the RxR-CE benchmark?.

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 9.0/10.

3 Results

4 papers retrieved. 9 claims extracted; 9 independently verified. Quality review score: 9.0/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
Embodied AI is widely recognized as a cornerstone of artificial general intelligence (AGI) because it involves controlli	✓	0.36
Vision-language-action (VLA) models have emerged to address language-conditioned robotic tasks in embodied AI by leverag	✓	0.41
The recent proliferation of VLAs necessitates a comprehensive survey to capture the rapidly evolving landscape.	✓	0.26
This work provides a detailed taxonomy of VLAs, organized into three major lines of research.	✓	0.24
The first line of research focuses on individual components of VLAs.	✓	0.17
The second line of research is dedicated to developing VLA-based control policies adept at predicting low-level actions.	✓	0.29
The third line of research comprises high-level task planners capable of decomposing long-horizon tasks into a sequence	✓	0.35
The survey provides an extensive summary of relevant resources, including datasets, simulators, and benchmarks.	✓	0.21
A curated repository associated with this survey is available at: https://github.com/yueen-ma/Awesome-VLA.	✓	0.29

References

- https://doi.org/10.48550/arxiv.2311.13549
- https://doi.org/10.59717/j.xinn-inform.2025.100015

- https://doi.org/10.1109/tnnls.2025.3650584