SOVEREIGN: What is the impact of VLA model scale (7B vs 13B) on object grounding accuracy and path completion rate in Lon

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

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 category of multimodal models-referred to as vision-language-action (VLA) models-has emerged to address language-conditioned robotic tasks in embodied AI by leveraging their distinct ability to generate actions. The recent proliferation of VLAs necessitates a comprehensive survey to capture the rapidly evolving landscape.

1 Introduction

Analysis of: A Survey on Vision–Language–Action Models for Embodied AI. Research goal: What is the impact of VLA model scale (7B vs 13B) on object grounding accuracy and path completion rate in LongNav-R1 on R2R-CE under varying instruction complexity?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

2 papers retrieved. 9 claims extracted, 9 verified. Tribunal: 9.0/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

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
A new category of multimodal models-referred to as vision-language-action (VLA) models-has emerged to address language-c	✓	0.47
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 focuses on individual components of VLAs.	✓	0.19
The second line is dedicated to developing VLA-based control policies adept at predicting low-level actions.	✓	0.30
The third line comprises high-level task planners capable of decomposing long-horizon tasks into a sequence of subtasks,	✓	0.36
We provide an extensive summary of relevant resources, including datasets, simulators, and benchmarks.	✓	0.23
A curated repository associated with this survey is available at: https://github.com/yueen-ma/Awesome-VLA.	✓	0.28

References

- https://doi.org/10.3348/kjr.2025.0599
- https://doi.org/10.1109/tnnls.2025.3650584