

# Scaling Laws of LAP Performance Across Unseen Robot Morphologies and Tasks

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does the performance of LAP scale with the number of training embodiments when evaluated on out-of-distribution robot morphologies, benchmarked using success rate metrics on unseen RoboStack tasks. 11 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: LAP: Language-Action Pre-Training Enables Zero-shot Cross-Embodiment Transfer. Research question: How does the performance of LAP scale with the number of training embodiments when evaluated on out-of-distribution robot morphologies, benchmarked using success rate metrics on unseen RoboStack tasks?.

## 2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

## 3 Results

14 papers retrieved. 11 claims extracted; 1 independently verified. Quality review score: 4.5/10.

## 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.

## 5 Extracted Claims

Claim	Verified	Confidence
LAP-3B achieves performance comparable to the $\pi$ 0.5-DROID on the seen embodiment.	×	0.07
Across three previously unseen embodiments and six real-world manipulation tasks, LAP-3B attains over 50% average zero-s	✓	0.27
LAP-3B delivers approximately a 2 $\times$ improvement over the strongest baselines.	×	0.09
All open-sourced VLAs collapse to zero success rate.	×	0.05
LAP is evaluated across four robot embodiments, ten real-world manipulation tasks, and the LIBERO simulation benchmark.	×	0.10
LAP-3B adopts a Mixture-of-Transformers architecture combining a LAP-trained VLM backbone with a lightweight flow-matchi	×	0.05
LAP-3B differs from $\pi$ 0.5 only in the action representation used to supervise the VLM (language-actions versus FAST token	×	0.07
The VLM backbone is optimized to predict structured language-actions using the cross-entropy loss in Eq. (1).	×	0.04
The action expert predicts continuous action chunks at:t+H via a flow-matching objective LFM.	×	0.02
The overall training objective is $L = LFM + \lambda LCE$ .	×	0.02
The VLM and action expert communicate solely through cross-attention.	×	0.03

## References

- <http://arxiv.org/abs/2505.05753v2>

- <http://arxiv.org/abs/2410.12124v2>
- <http://arxiv.org/abs/2602.10556v2>