Multimodal Pretraining and Scale Effects on Zero-Shot Reasoning in 13B vs 7B VLA Models

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: Do 13B VLA models with multimodal pretraining demonstrate better zero-shot reasoning capabilities on the MM-ReAct benchmark compared to 7B models when evaluated using Exact Match accuracy. Web-crawled pretraining datasets underlie the impressive "zero-shot" evaluation performance of multimodal models, such as CLIP for classification/retrieval and Stable-Diffusion for image generation. However, it is unclear how meaningful the notion of "zero-shot" generalization. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance. Research question: Do 13B VLA models with multimodal pretraining demonstrate better zero-shot reasoning capabilities on the MM-ReAct benchmark compared to 7B models when evaluated using Exact Match accuracy?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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/2404.04125v3
- http://arxiv.org/abs/2603.14523v1
- http://arxiv.org/abs/2411.15127v3