Multimodal Models Pre-Trained on Visual Genome Enhance Adversarial Robustness in Visual Reasoning

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: Do multimodal models pre-trained on Visual Genome exhibit improved robustness against adversarial visual perturbations in visual reasoning tasks compared to models trained on sparse image-text pairs. 6 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.9/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Foundation models for generalist medical artificial intelligence. Research question: Do multimodal models pre-trained on Visual Genome exhibit improved robustness against adversarial visual perturbations in visual reasoning tasks compared to models trained on sparse image-text pairs?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.9/10.

3 Results

10 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 7.9/10.

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
The exceptionally rapid development of highly flexible, reusable artificial intelligence (AI) models is likely to usher	✓	0.33
GMAI models will be capable of carrying out a diverse set of tasks using very little or no task-specific labelled data.	✓	0.31
Built through self-supervision on large, diverse datasets, GMAI will flexibly interpret different combinations of medica	✓	0.42
Models will in turn produce expressive outputs such as free-text explanations, spoken recommendations or image annotatio	✓	0.34
GMAI-enabled applications will challenge current strategies for regulating and validating AI devices for medicine.	✓	0.27
GMAI-enabled applications will shift practices associated with the collection of large medical datasets.	✓	0.26

References

- https://doi.org/10.1109/jproc.2021.3054390
- https://doi.org/10.1038/s41586-023-05881-4
- https://doi.org/10.3390/bioengineering11040337