

# Semantic Alignment and Zero-Shot Accuracy in Synthetic Image-Text Pairs

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the semantic alignment score of synthetic image-text pairs generated by multimodal models correlate with zero-shot accuracy degradation on the VQAv2 benchmark. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Semantic Object Accuracy for Generative Text-to-Image Synthesis. Research question: How does the semantic alignment score of synthetic image-text pairs generated by multimodal models correlate with zero-shot accuracy degradation on the VQAv2 benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.0/10.

## 3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.0/10.

## 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.

## References

- <http://arxiv.org/abs/1910.13321v2>
- <http://arxiv.org/abs/2603.03437v1>
- <http://arxiv.org/abs/2510.03260v1>