

# GOLIATH-Style Oversampling and Adversarial Training for Multimodal Alignment Robustness

Assignee Research

June 8, 2026

## Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the impact of GOLIATH-style oversampling versus adversarial training on the alignment robustness of multimodal models under extreme class imbalance. 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: Stop Oversampling for Class Imbalance Learning: A Critical Review. Research question: What is the impact of GOLIATH-style oversampling versus adversarial training on the alignment robustness of multimodal models under extreme class imbalance?.

## 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 6.0/10.

## 3 Results

12 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/2605.14067v1>
- <http://arxiv.org/abs/2202.03579v2>
- <http://arxiv.org/abs/2308.02966v1>