

Cross-Domain Adversarial Transfer and Alignment Stability in Multimodal Models

Assignee Research

June 3, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: How does cross-domain adversarial transfer between image and text modalities affect the alignment stability of multimodal models under attack. 9 claims were extracted from source literature; 9 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Text Data Augmentation for Deep Learning. Research question: How does cross-domain adversarial transfer between image and text modalities affect the alignment stability of multimodal models under attack?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.8/10.

3 Results

8 papers retrieved. 9 claims extracted; 9 independently verified. Quality review score: 8.8/10.

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
Natural Language Processing (NLP) is one of the most captivating applications of Deep Learning.	✓	0.25
Data Augmentation training strategy can aid in the development of NLP.	✓	0.24
The major motifs of Data Augmentation include strengthening local decision boundaries, brute force training, causality a	✓	0.36
Deep Learning generally struggles with the measurement of generalization and characterization of overfitting.	✓	0.25
Augmentations can construct test sets for generalization.	✓	0.20
NLP is at an early stage in applying Data Augmentation compared to Computer Vision.	✓	0.29
There are key differences and promising ideas that have yet to be tested in NLP regarding Data Augmentation.	✓	0.24
Tools that facilitate Data Augmentation include the use of consistency regularization, controllers, and offline and onli	✓	0.29
Interesting topics around Data Augmentation in NLP include task-specific augmentations, the use of prior knowledge in se	✓	0.46

References

- <https://doi.org/10.1109/tnnls.2021.3070843>
- <https://doi.org/10.1186/s40537-021-00492-0>
- <https://doi.org/10.4230/lipics.cosit.2024.11>