Counterfactual Text Augmentation for Robustness in Multimodal VQA Models

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: What is the effect of counterfactual text augmentation on the robustness of multimodal models against adversarial perturbations in VQA tasks. Deep neural networks (DNNs) are an indispensable machine learning tool despite the difficulty of diagnosing what aspects of a model's input drive its decisions. In countless real-world domains, from legislation and law enforcement to healthcare, such diagnosis is essential to. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Explainable Deep Learning: A Field Guide for the Uninitiated. Research question: What is the effect of counterfactual text augmentation on the robustness of multimodal models against adversarial perturbations in VQA tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

3 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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

- https://doi.org/10.1613/jair.1.13200
- https://doi.org/10.1007/s12559-023-10179-8
- https://doi.org/10.1016/j.ejrad.2023.110786