Comparative Analysis of Clustering and Projection Debiasing Methods for Gender Bias in Contextualized Embeddings

Assignee Research

June 11, 2026

Abstract

Taking an interdisciplinary approach to surveying issues around gender bias in textual and visual AI, we present literature on gender bias detection and mitigation in NLP, CV, as well as combined visual-linguistic models. We identify conceptual parallels between these strands of research as well as how methodologies were adapted cross-disciplinary from NLP to CV. We also find that there is a growing awareness for theoretical frameworks from the social sciences around gender in NLP that could be beneficial for aligning bias analytics in CV with human values and conceptualising gender beyond the

1 Introduction

This paper examines: Gender Bias in Natural Language Processing and Computer Vision: A Comparative Survey. Research question: Do clustering-based debiasing methods for contextualized embeddings demonstrate better downstream task performance on gender-bias benchmarks like WinoBias or CrowS-Pairs compared to projection-based methods?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.0/10.

3 Results

4 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 8.0/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
Gender bias is present in both natural language processing (NLP) and computer vision (CV) systems.	✓	0.25
There are conceptual parallels between gender bias detection and mitigation in NLP and CV.	✓	0.33
Methodologies for addressing gender bias have been adapted from NLP to CV.	✓	0.20
There is a growing awareness in NLP for theoretical frameworks from the social sciences around gender.	✓	0.26
Theoretical frameworks from the social sciences could be beneficial for aligning bias analytics in CV with human values.	✓	0.36
Conceptualising gender beyond the binary categories of male/female is important for addressing gender bias in AI systems	✓	0.26

References

- https://doi.org/10.18653/v1/2025.trustnlp-main.31
- https://doi.org/10.1145/3700438
- https://doi.org/10.3390/fi17080340