

Lion Optimizer’s Impact on ModernBERT Cross-Encoder Robustness Against Domain Shift

Assignee Research

June 11, 2026

Abstract

Language models have revolutionized natural language processing, becoming an integral part of many applications. However, these models often exhibit societal biases embedded in their training data, raising concerns about their fairness and ethical deployment. Measuring these biases usually requires creating datasets with time-consuming human annotation, which is costly and hard to expand. To address this challenge, we propose a data curation framework and CineBias, a novel dataset of 1,012 stereotypical sentence pairs covering seven bias categories, extracted from Hollywood movie subtitles wit

1 Introduction

This paper examines: Keelemudelite sotsiaalse kallutatuse uurimine filmisubtiitrite põhjal. Research question: Does the Lion optimizer improve the robustness of ModernBERT cross-encoders against domain shift when evaluated on out-of-distribution retrieval benchmarks beyond MS MARCO?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.7/10.

3 Results

9 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 8.7/10.

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.

5 Extracted Claims

| Claim | Verified | Confidence |
|--|----------|------------|
| Language models often exhibit societal biases embedded in their training data. | ✓ | 0.26 |
| Measuring societal biases in language models usually requires creating datasets with time-consuming human annotation, wh | ✓ | 0.34 |
| A novel dataset of 1,012 stereotypical sentence pairs covering seven bias categories, named CineBias, has been extracted | ✓ | 0.39 |
| The language models BERT, RoBERTa, and ModernBERT were evaluated using the CrowS-Pairs Score (CPS) on CineBias. | ✓ | 0.29 |
| BERT achieved a bias level of 61.2% CPS on the CineBias dataset. | ✓ | 0.15 |
| CineBias provides a scalable way to measure bias in language models. | ✓ | 0.25 |
| The applicability of CineBias to low-resource languages was demonstrated with an Estonian case study. | ✓ | 0.19 |

References

- <https://openalex.org/W7148199498>
- <https://openalex.org/W7113234157>
- <https://doi.org/10.18653/v1/2023.emnlp-main.623>