

DPO-Aligned Response Quality and Inference Latency Variance in Low- and High-Resource Languages for Hate Speech Mitigation

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the correlation between DPO-aligned response quality scores and inference latency variance across low-resource versus high-resource languages in hate speech mitigation tasks. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: One to rule them all: Towards Joint Indic Language Hate Speech Detection. Research question: What is the correlation between DPO-aligned response quality scores and inference latency variance across low-resource versus high-resource languages in hate speech mitigation tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

10 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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.

References

- <http://arxiv.org/abs/2412.16365v1>
- <http://arxiv.org/abs/2101.03207v1>
- <http://arxiv.org/abs/2109.13711v1>