Direct Preference Optimization Impact on Multilingual Counter-Speech Quality and Latency Trade-offs

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the effect of LLM alignment via DPO on the trade-off between response quality (evaluated via human judgment metrics) and inference latency in multilingual counter-speech generation tasks. 17 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Northeastern Uni at Multilingual Counterspeech Generation: Enhancing Counter Speech Generation with LLM Alignment through Direct Preference Optimization. Research question: What is the effect of LLM alignment via DPO on the trade-off between response quality (evaluated via human judgment metrics) and inference latency in multilingual counter-speech generation tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

13 papers retrieved. 17 claims extracted; 1 independently verified. Quality review score: 4.2/10.

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
The model significantly outperforms SFT baselines on CS benchmarks while scaling effectively to multiple languages.	✓	0.23
All training processes were executed on a single 32 GB V-100 GPU.	×	0.02
Supervised fine-tuning was applied using the Llama3 basic and instruct models with default parameters and LoRA fine-tuni	×	0.08
Default parameters included a batch size of 4, combining gradients over 4 steps, and weight decay of 0.01.	×	0.00
For LoRA, the rank (r) was set to 16, the scaling factor (alpha) to 16, and a dropout of 0 was applied to the low-rank l	×	0.02
The training dataset consisted of only 1,500 lines, necessitating a higher number of epochs to sufficiently train the SF	×	0.04
The maximum sequence length was set to 640 to prevent excessively long outputs.	×	0.01
The Adam optimizer was employed with a learning rate of 2e-4, conducting training for 500 epochs for each model.	×	0.02
The entire training process spanned approximately 70 hours.	×	0.04
Checkpoints at 150 epochs for the Llama3 basic model and 200 epochs for the Llama3 instruct model were selected.	×	0.01
Training was extended on the DPO dataset based on the SFT checkpoint with a learning rate of 5e-4 for an additional 80 e	×	0.04
Improvements were observed in the basic model, while the instruct model showed no significant improvements.	×	0.03
Metrics used include AVG BLEU-2, BERTScore, JudgeLM, and AVG ROUGE-L.	×	0.02
Run3, the DPO-aligned Llama3 base model, outperforms the other runs across all metrics, followed by run2 (SFT Llama3 ins	×	0.06
The superior performance of run3 reinforces the efficacy of Direct Preference Optimization (DPO) for improving text gene	×	0.13
Standard supervised fine-tuning (SFT) is effective in generating coherent text but often fails to directly challenge and	×	0.08
The integration of metrics such as BLEU-2, BERTScore, JudgeLM, and ROUGE-L assesses the quality of the generated outputs	×	0.04

References

- http://arxiv.org/abs/2412.15453v1
- http://arxiv.org/abs/2508.04149v2
- http://arxiv.org/abs/2407.14477v4