Scalability and Performance of DPO-Aligned vs SFT-Only Models in Low-Resource Counter-Speech Generation

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the scalability of DPO-aligned models compare to SFT-only models when generating counter-speech in low-resource languages, measured by response quality metrics such as BLEU or ROUGE on the. 19 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. 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: How does the scalability of DPO-aligned models compare to SFT-only models when generating counter-speech in low-resource languages, measured by response quality metrics such as BLEU or ROUGE on the HateSpeech-Counterspeech dataset?.

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 3.5/10.

3 Results

10 papers retrieved. 19 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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 proposed model significantly outperforms SFT baselines on counterspeech (CS) benchmarks.	×	0.14
The model scales effectively to multiple languages.	×	0.10
Model supervision and alignment were performed in English.	×	0.12
All training processes were executed on a single 32 GB V-100 GPU.	×	0.02
Supervised fine-tuning (SFT) was applied using Llama3 basic and instruct models with LoRA techniques.	×	0.12
SFT default parameters included a batch size of 4, gradient accumulation over 4 steps, and a weight decay of 0.01.	×	0.01
LoRA configuration used a rank (r) of 16, a scaling factor (alpha) of 16, and a dropout of 0 targeting attention layers.	×	0.02
The training dataset consisted of 1,500 lines.	×	0.02
The maximum sequence length was set to 640 tokens.	×	0.02
The Adam optimizer was used with a learning rate of 2e-4 for SFT training.	×	0.04
SFT training was conducted for 500 epochs for each model.	×	0.05
The entire SFT training process spanned approximately 70 hours.	×	0.02
The selected checkpoint for the Llama3 basic model (run1) was at 150 epochs.	×	0.01
The selected checkpoint for the Llama3 instruct model (run2) was at 200 epochs.	×	0.01
DPO training used a learning rate of 5e-4 and continued for an additional 80 epochs.	×	0.05
Evaluation metrics included AVG BLEU-2, BERTScore, JudgeLM, and AVG ROUGE-L.	×	0.05
Run3 (DPO-aligned Llama3 base model) outperformed run2 and run1 across all evaluation metrics.	×	0.08
Run2 corresponds to the SFT Llama3 instruct model.	×	0.04
Run1 corresponds to the SFT Llama3 base model.	×	0.06

References

- http://arxiv.org/abs/2403.14938v1
- http://arxiv.org/abs/2412.15453v1
- http://arxiv.org/abs/2110.06500v2