Impact of OpenPangu-7B-MLA Human Preference Alignment on EchoMind Classification Versus Prosody-Exclusive Models Under Edge

Assignee Research

June 12, 2026

Abstract

Fine-grained control over large language models (LLMs) remains a significant challenge, hindering their adaptability to diverse user needs. While Reinforcement Learning from Human Feedback (RLHF) shows promise in aligning LLMs, its reliance on scalar rewards often limits its ability to capture diverse user preferences in real-world applications. To address this limitation, we introduce the Directional Preference Alignment (DPA) framework. Unlike the scalar-reward RLHF, DPA incorporates multi-objective reward modeling to represent diverse preference profiles. Additionally, DPA models user prefe

1 Introduction

This paper examines: Arithmetic Control of LLMs for Diverse User Preferences: Directional Preference Alignment with Multi-Objective Rewards. Research question: How does the alignment of OpenPangu-7B-MLA with human preferences affect its performance in EchoMind classification tasks on edge devices, compared to prosody-exclusive models, when evaluated using metrics like accuracy and response coherence under fixed hardware constraints?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.4/10.

3 Results

16 papers retrieved. 13 claims extracted; 10 independently verified. Quality review score: 7.4/10.

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
Figure 2 (Right) shows that the preferences of User-1, User-2, and User-3 can be accurately represented by specifying th	✓	0.29
Directional Preference Alignment (DPA) can alleviate the problem of misspecification in RLHF.	✓	0.23
The proposed approach involves Multi-Objective Rewards, which include learning with multiple different preference target	✓	0.16
Directional Preference Alignment encodes user preferences as unit vectors for preference-aware LLM alignment.	✓	0.27
Existing popular RLHF frameworks have limited capacity for capturing real-world complicated human preferences.	✓	0.24
Existing popular RLHF frameworks lack adaptability for user-dependent preferences.	✓	0.15
Directional Preference Alignment (DPA) allows a single LLM to accommodate users with varying preferences.	✓	0.24
The study considers both helpfulness and verbosity rewards.	×	0.08
The Mistral-7B model was aligned using the proposed DPA method.	×	0.11
Empirical evaluations show that DPA offers effective arithmetic control over the trade-off between helpfulness and verbo	✓	0.23
Empirical evaluations show that DPA maintains competitive performance with DPO (Rafailov et al., 2023).	✓	0.20
The Linear Scalarization method uses the formula R = v1 * helpfulness + v2 * verbosity.	✓	0.15
In the described Linear Scalarization example, the values v1 = 0.8 and v2 = 0.6 are used.	×	0.09

References

- http://arxiv.org/abs/2201.08780v2
- http://arxiv.org/abs/2310.04793v2
- http://arxiv.org/abs/2402.18571v3