Alignment and Inference Efficiency Trade-offs in Cross-Domain Fine-Tuned LLMs on DS-1000

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: Does the alignment of cross-domain fine-tuned LLMs with human preferences (e.g., via RLHF) influence their inference efficiency on the DS-1000 benchmark, as measured by tokens per second throughput. 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. 10 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Arithmetic Control of LLMs for Diverse User Preferences: Directional Preference Alignment with Multi-Objective Rewards. Research question: Does the alignment of cross-domain fine-tuned LLMs with human preferences (e.g., via RLHF) influence their inference efficiency on the DS-1000 benchmark, as measured by tokens per second throughput and functional correctness trade-offs?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

14 papers retrieved. 10 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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
Figure 2 (Right) shows that the preferences of User-1, User-2, and User-3 can be accurately represented by specifying th	×	0.07
Directional Preference Alignment (DPA) can alleviate the problem of misspecification in RLHF.	×	0.13
Existing popular RLHF frameworks have limited capacity for capturing real-world complicated human preference.	×	0.07
Existing popular RLHF frameworks lack adaptability for user-dependent preference.	×	0.10
Directional Preference Alignment (DPA) allows a single LLM to accommodate users with varying preferences.	×	0.13
The study aligns the Mistral-7B model using the proposed DPA method.	×	0.08
Empirical evaluations show that DPA offers effective arithmetic control over the trade-off between helpfulness and verbo	✓	0.22
Empirical evaluations show that DPA maintains competitive performance with DPO (Rafailov et al., 2023).	×	0.05
The Linear Scalarization method uses a reward function R = v1 * helpfulness + v2 * verbosity.	×	0.07
In the described Linear Scalarization example, the parameters are set to v1 = 0.8 and v2 = 0.6.	×	0.03

References

- http://arxiv.org/abs/2602.00426v1
- http://arxiv.org/abs/2402.18571v3
- http://arxiv.org/abs/2304.05302v3