

# Directional Preference Alignment for Stable Python Code Generation Under Input Perturbations

Assignee Research

May 31, 2026

## Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: Can Directional Preference Alignment maintain higher inference stability than RLHF under edge-case input perturbations in Python code generation evaluations. 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; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. 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: Can Directional Preference Alignment maintain higher inference stability than RLHF under edge-case input perturbations in Python code generation evaluations?.

## 2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

### 3 Results

8 papers retrieved. 10 claims extracted; 2 independently verified. Quality review score: 5.2/10.

### 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 proposed Directional Preference Alignment (DPA) approach allows a single LLM to accommodate users with varying preferences.	×	0.13
DPA offers effective arithmetic control over the trade-off between helpfulness and verbosity.	✓	0.22
DPA maintains competitive performance with DPO (Rafailov et al., 2023).	×	0.04
The preferences of User-1, User-2, and User-3 can be accurately represented by specifying the preference vector in the 2D space.	×	0.08
DPA can alleviate the problem of misspecification in RLHF.	×	0.04
Existing popular RLHF frameworks have limited capacity for capturing the real-world complicated human preference.	×	0.09
Existing popular RLHF frameworks lack adaptability for user-dependent preference.	×	0.11
The linear scalarization method uses $R = v_1 \cdot \text{helpfulness} + v_2 \cdot \text{verbosity}$ with $v_1 = 0.8$ and $v_2 = 0.6$ .	×	0.04
The empirical evaluations show that DPA offers effective arithmetic control over the trade-off between helpfulness and verbosity.	✓	0.19
The empirical evaluations show that DPA maintains competitive performance with DPO.	×	0.07

## References

- <http://arxiv.org/abs/2407.14477v4>
- <http://arxiv.org/abs/2402.18571v3>
- <http://arxiv.org/abs/2312.11456v4>