Directional Preference Alignment with Multi-Objective Rewards Enhances Robustness to Distributional Shifts in User Preferences

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How does Directional Preference Alignment with multi-objective rewards improve robustness to distributional shifts in user preferences compared to scalar-reward RLHF, as measured by. 11 claims were extracted from source literature; 11 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Bone Soups: A Seek-and-Soup Model Merging Approach for Controllable Multi-Objective Generation. Research question: How does Directional Preference Alignment with multi-objective rewards improve robustness to distributional shifts in user preferences compared to scalar-reward RLHF, as measured by Helpfulness-Harmlessness scores on out-of-distribution evaluations in PaLM 2 variants?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.2/10.

3 Results

4 papers retrieved. 11 claims extracted; 11 independently verified. Quality review score: 8.2/10.

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
User information needs are often highly diverse and varied.	✓	0.20
A key challenge in current research is how to achieve controllable multi-objective generation while enabling rapid adapt	✓	0.37
Existing solutions, such as Rewarded Soup, focus on merging language models individually tuned on single objectives.	✓	0.29
These approaches face limitations in achieving optimal performance due to their disregard for the impacts of competing o	✓	0.29
Bone Soup is a novel model merging approach that first seeks a series of backbone models by considering the impacts of m	✓	0.44
Bone Soup begins by training multiple backbone models for different objectives using multi-objective reinforcement learn	✓	0.36
Each backbone model is guided by a combination of backbone reward signals.	✓	0.23
The backbone rewards are crafted by combining standard reward functions into basis vectors, which can then be modified t	✓	0.25
Bone Soup leverages a symmetric circulant matrix mapping to generate the merging coefficients.	✓	0.28
The merging coefficients are used to merge the backbone models according to user preferences.	✓	0.30
Extensive experimental results demonstrate that Bone Soup exhibits strong controllability and Pareto optimality in contr	✓	0.37

References

- https://doi.org/10.48550/arxiv.2405.07863
- https://doi.org/10.48550/arxiv.2402.02030
- https://doi.org/10.18653/v1/2025.acl-long.1322