Multimodal Data Diversity and PaLM-E Alignment with Human Preferences in Visual Problem-Solving

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: To what extent does increasing the diversity of multimodal training data improve PaLM-E's alignment scores with human preferences in visual problem-solving tasks. 17 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.1/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Aligning Multimodal LLM with Human Preference: A Survey. Research question: To what extent does increasing the diversity of multimodal training data improve PaLM-E's alignment scores with human preferences in visual problem-solving tasks?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.1/10.

3 Results

13 papers retrieved. 17 claims extracted; 0 independently verified. Quality review score: 3.1/10.

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
Fact-RLHF is the first multimodal RLHF algorithm.	×	0.04
Fact-RLHF utilizes 10K human-labeled samples for the reward model.	×	0.08
Fact-RLHF utilizes 50K hold-out data.	×	0.02
The loss function of Fact-RLHF integrates a per-token KL penalty, factual information to calibrate judgments, and correc	×	0.01
DDPO assigns higher weights to corrected data in its loss function compared to standard DPO.	×	0.04
DDPO uses 1.4K manually refined samples.	×	0.00
In the DDPO dataset, objects account for 41.2% of hallucination types.	×	0.02
In the DDPO dataset, positions account for 20.3% of hallucination types.	×	0.02
In the DDPO dataset, numbers account for 16.5% of hallucination types.	×	0.02
In the DDPO dataset, attributes account for 10% of hallucination types.	×	0.02
In the DDPO dataset, actions account for 5.3% of hallucination types.	×	0.02
In the DDPO dataset, other hallucination types account for 6.8%.	×	0.02
FDPO reuses InstructBLIP.	×	0.03
This survey is the first to specifically focus on the alignment of MLLMs.	×	0.08
Existing surveys focus on the alignment of AI but none specifically address the alignment of MLLMs.	×	0.06
MLLM alignment algorithms are developed to address the issue of hallucinations in multimodal systems.	×	0.09
Recent research shows that MLLM alignment algorithms improve conversational capabilities, reasoning abilities, and other	×	0.07

References

- `http://arxiv.org/abs/2604.00086v1`

- http://arxiv.org/abs/2503.14504v2
- http://arxiv.org/abs/2407.14477v4