Preference Alignment Training Effects on Multimodal Clinical Model Triage Consistency

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the impact of preference alignment training on multimodal clinical models when evaluated against human expert rankings for triage decision consistency. 13 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: Aligning Multimodal LLM with Human Preference: A Survey. Research question: What is the impact of preference alignment training on multimodal clinical models when evaluated against human expert rankings for triage decision consistency?.

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 4.5/10.

3 Results

13 papers retrieved. 13 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
MME-RealWorld, MMStar, MMBench, MMT-Bench, BLINK, MathVista, SQA3D, MMMU, MVBench, Mantis-Instruct are benchmarks for Ge	×	0.03
Object HalBench, VideoHallucer, VALOR-Eval, POPE, HaELM, OpenCHAIR, GAVIE, AMBER, Mementos, MMHal-Bench, VLind-Bench, M-	×	0.01
AdvDiffVLM, RTVLM, VLGuard, MultiTrust, VLLM-safety-bench, MOSSBench, MM-RLHF-SafetyBench are benchmarks for Safety eval	×	0.02
Q-Bench, LLVisionQA, LLDescribe, LLaVA-Bench-Wilder, LiveBench, Vibe-Eval are benchmarks for Conversation evaluation.	×	0.02
M-RewardBench, VL-RewardBench, RewardBench, MJ-Bench, MLLM-as-a-Judge, MM-RLHF-RewardBench are benchmarks for Reward Mod	×	0.03
Arena-Hard, AlpacaEval-V2, AlignBench, MM-AlignBench are benchmarks for Alignment evaluation.	×	0.02
Fact-RLHF is the first multimodal RLHF algorithm, utilizing 10K human-labeled samples for the reward model and 50K hold-	×	0.05
DDPO assigns higher weights to corrected data in its loss function compared to standard DPO.	×	0.04
DDPO uses 1.4K manually refined samples covering hallucination types such as objects (41.2%), positions (20.3%), numbers	×	0.02
FDPO reuses InstructBLIP's existing data.	×	0.03
The creation of alignment datasets involves three core factors: data sources, model responses, and preference annotation	✓	0.19
Most alignment algorithms are designed for specific tasks such as addressing hallucinations, ensuring safety, and improv	×	0.11
This survey is the first to specifically focus on the alignment of MLLMs.	×	0.07

References

- http://arxiv.org/abs/2312.11456v4
- http://arxiv.org/abs/2503.14504v2
- http://arxiv.org/abs/2407.14477v4