

Multi-Modal Lightweight Transformers vs. Text-Only Models in Code Generation and Reasoning Benchmarks

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How do multi-modal lightweight Transformers perform relative to text-only models on mixed code-generation and reasoning benchmarks (e.g., MBPP + MMLU) when evaluated for alignment with human. Large language models (LLMs) can handle a wide variety of general tasks with simple prompts, without the need for task-specific training. Multimodal Large Language Models (MLLMs), built upon LLMs, have demonstrated impressive potential in tackling complex tasks involving visual. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. 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: How do multi-modal lightweight Transformers perform relative to text-only models on mixed code-generation and reasoning benchmarks (e.g., MBPP + MMLU) when evaluated for alignment with human preferences?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

12 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 3.8/10.

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 ev	×	0.03
Object HalBench, VideoHalluciner, 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 evaluating	×	0.03
Q-Bench, LLVisionQA, LLDescribe, LLaVA-Bench-Wilder, LiveBench, Vibe-Eval are benchmarks for evaluating conversation in	×	0.02
M-RewardBench, VL-RewardBench, RewardBench, MJ-Bench, MLLM-as-a-Judge, MM-RLHF-RewardBench are benchmarks for evalu-	×	0.04
ating		
Arena-Hard, AlpacaEval-V2, AlignBench, MM-AlignBench are benchmarks for evaluating align-	×	0.03
ment in multimodal models.		
Fact-RLHF is the first multimodal RLHF algo-	×	0.06
rithm, utilizing 10K human-labeled samples for the reward model and 50K hold-		
DDPO assigns higher weights to corrected data	×	0.04
in its loss function compared to standard DPO.		
DDPO uses 1.4K manually refined samples	×	0.01
covering hallucination types such as objects		
(41.2%), positions (20.3%), numbers		
FDPO reuses InstructBLIP’s existing data.	×	0.04
The creation of alignment datasets involves three	✓	0.18
core factors: data sources, model responses, and		
preference annotation		
Most alignment algorithms are designed for spe-	×	0.11
cific tasks such as addressing hallucinations, en-		
suring safety, and improv		
This survey is the first to specifically focus on	×	0.08
the alignment of MLLMs.		

References

- <http://arxiv.org/abs/2503.14504v2>
- <http://arxiv.org/abs/2505.20977v3>
- <http://arxiv.org/abs/2402.18571v3>