

# Cross-Domain Multimodal Alignment Enhances Contextual Safety in Llama-3.1-8B on MTMCS-Bench

Assignee Research

June 7, 2026

## Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: Can cross-domain multimodal alignment techniques improve the contextual safety scores of models like Llama-3.1-8B on MTMCS-Bench compared to models optimized solely on single-turn benchmarks. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: Dynamic Multimodal Fusion via Meta-Learning Towards Micro-Video Recommendation. Research question: Can cross-domain multimodal alignment techniques improve the contextual safety scores of models like Llama-3.1-8B on MTMCS-Bench compared to models optimized solely on single-turn benchmarks?.

## 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 3.5/10.

## 3 Results

4 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 3.5/10.

## 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
VBPR obtains poor performance on three datasets.	×	0.04
MetaMMF_MF consistently outperforms VBPR across all cases.	×	0.02
GCN-based methods show better performance on micro-video recommendation compared to MF-based methods.	×	0.15
The dataset MovieLens has 55,485 users, 5,986 items, and 1,239,508 interactions with a density of 0.37%.	×	0.01
The dataset TikTok has 36,656 users, 76,085 items, and 726,065 interactions with a density of 0.03%.	×	0.01
The dataset Kwai has 7,010 users, 86,483 items, and 298,492 interactions with a density of 0.05%.	×	0.01
MetaMMF is model-agnostic and can be used as the early step prior to other representation learning functions, such as MF	×	0.10
MetaMMF dynamically learns the parameter of a neural layer by processing the meta information.	×	0.12

## References

- <http://arxiv.org/abs/2408.07303v2>
- <http://arxiv.org/abs/2501.07110v1>
- <http://arxiv.org/abs/2412.17297v1>