SOVEREIGN: To what extent does modality imbalance affect the accuracy and routing stability of multimodal language models

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

The rise of Multimodal Large Language Models (MLLMs) has significantly advanced the capabilities of AI systems to understand and generate content across diverse modalities such as text, images, audio, video, and sensory data. By leveraging the reasoning prowess of Large Language Models (LLMs), MLLMs unify multiple input formats into a coherent framework, enabling unprecedented performance in multimodal tasks. This survey provides a comprehensive overview of the architectural innovations, training paradigms, data resources, and evaluation benchmarks that have shaped the evolution of MLLMs. We r

1 Introduction

Analysis of: Multimodal Large Language Models: Developments and Directions. Research goal: To what extent does modality imbalance affect the accuracy and routing stability of multimodal language models as measured by performance on MMBench and SEED-Bench evaluation suites?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

1 papers retrieved. 10 claims extracted, 10 verified. Tribunal: 7.8/10 → APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
Multimodal Large Language Models (MLLMs) leverage the reasoning prowess of Large Language Models (LLMs) to unify multipl	✓	0.35
MLLMs enable unprecedented performance in multimodal tasks.	✓	0.17
The survey provides a comprehensive overview of architectural innovations, training paradigms, data resources, and evalu	✓	0.31
The survey reviews foundational and emerging models, tracing contributions from both academia and industry.	✓	0.19
Recent efforts have expanded modality coverage, improved cross-lingual support, and enabled interactive and embodied AI.	✓	0.18
Key research themes discussed include multimodal hallucination, multimodal in-context learning (M-ICL), chain-of-thought	✓	0.45
Recent benchmarks such as MMBench, SEED-Bench, and MathVista are introduced in the survey.	✓	0.17
Trends in aligning models with human preferences via techniques like Reinforcement Learning with Human Feedback (RLHF) a	✓	0.27
Persistent challenges include scalability, multimodal alignment, interpretability, safety, and fairness.	✓	0.20
Open research directions are identified that can guide future exploration.	✓	0.16

References

- https://www.semanticscholar.org/paper/3e64fd5d817f97110c6df7e19ce61a87602b999b