What is the impact of modality dropout on the alignment accuracy of large multimodal models versus unimodal la

Assignee Research

June 10, 2026

Abstract

Reasoning lies at the heart of intelligence, shaping the ability to make decisions, draw conclusions, and generalize across domains. In artificial intelligence, as systems increasingly operate in open, uncertain, and multimodal environments, reasoning becomes essential for enabling robust and adaptive behavior. Large Multimodal Reasoning Models (LMRMs) have emerged as a promising paradigm, integrating modalities such as text, images, audio, and video to support complex reasoning capabilities and aiming to achieve comprehensive perception, precise understanding, and deep reasoning. As research

1 Introduction

This paper examines: Perception, Reason, Think, and Plan: A Survey on Large Multimodal Reasoning Models. Research question: What is the impact of modality dropout on the alignment accuracy of large multimodal models versus unimodal language models in zero-shot reasoning tasks?.

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. 0 claims extracted; 0 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.

References

- http://arxiv.org/abs/2308.10783v2
- http://arxiv.org/abs/2306.05540v1
- http://arxiv.org/abs/2505.04921v2