

How does the zero-shot cross-domain retrieval performance of MMICL compare to specialized multimodal models on

Assignee Research

May 29, 2026

Abstract

Strong Artificial Intelligence (Strong AI) or Artificial General Intelligence (AGI) with abstract reasoning ability is the goal of next-generation AI. Recent advancements in Large Language Models (LLMs), along with the emerging field of Multimodal Large Language Models (MLLMs), have demonstrated impressive capabilities across a wide range of multimodal tasks and applications. Particularly, various MLLMs, each with distinct model architectures, training data, and training stages, have been evaluated across a broad range of MLLM benchmarks. These studies have, to varying degrees, revealed differ

1 Introduction

This paper examines: Exploring the Reasoning Abilities of Multimodal Large Language Models (MLLMs): A Comprehensive Survey on Emerging Trends in Multimodal Reasoning. Research question: How does the zero-shot cross-domain retrieval performance of MMICL compare to specialized multimodal models on TextCaps when evaluated using precision@K and mean average precision (mAP) metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 6 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.5/10.

3 Results

6 papers retrieved. 10 claims extracted; 9 independently verified. Quality review score: 8.5/10.

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
Strong Artificial Intelligence (Strong AI) or Artificial General Intelligence (AGI) with abstract reasoning ability is t	✓	0.38
Recent advancements in Large Language Models (LLMs), along with the emerging field of Multimodal Large Language Models (✓	0.46
Various MLLMs, each with distinct model architectures, training data, and training stages, have been evaluated across a	✓	0.34
These studies have, to varying degrees, revealed different aspects of the current capabilities of MLLMs.	✓	0.28
The reasoning abilities of MLLMs have not been systematically investigated.	✓	0.25
This survey comprehensively reviews the existing evaluation protocols of multimodal reasoning.	✓	0.22
This survey categorizes and illustrates the frontiers of MLLMs.	×	0.09
This survey introduces recent trends in applications of MLLMs on reasoning-intensive tasks.	✓	0.27
This survey discusses current practices and future directions in multimodal reasoning.	✓	0.21
This survey establishes a solid base and sheds light on the important topic of multimodal reasoning.	✓	0.32

References

- <https://doi.org/10.48550/arxiv.2401.06805>
- <https://doi.org/10.48550/arxiv.2407.03320>

- <https://doi.org/10.48550/arxiv.2311.07594>