

# Do multimodal models fine-tuned on causal-augmented data outperform those fine-tuned on synthetic minority ove

Assignee Research

June 10, 2026

## Abstract

Creativity is a fundamental aspect of intelligence, involving the ability to generate novel and appropriate solutions across diverse contexts. While Large Language Models (LLMs) have been extensively evaluated for their creative capabilities, the assessment of Multimodal Large Language Models (MLLMs) in this domain remains largely unexplored. To address this gap, we introduce Creation-MMBench, a multimodal benchmark specifically designed to evaluate the creative capabilities of MLLMs in real-world, image-based tasks. The benchmark comprises 765 test cases spanning 51 fine-grained tasks. To ens

## 1 Introduction

This paper examines: Creation-MMBench: Assessing Context-Aware Creative Intelligence in MLLM. Research question: Do multimodal models fine-tuned on causal-augmented data outperform those fine-tuned on synthetic minority oversampled data in out-of-distribution generalization on MMBench, as measured by accuracy on unseen subsets?.

## 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 6.3/10.

## 3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.3/10.

## 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/2503.14478v2>
- <http://arxiv.org/abs/2110.06500v2>
- <http://arxiv.org/abs/2308.10783v2>