

Scaling MMC-Instruction Data for Robustness to Chart Type Distribution Shifts

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the impact of increasing the scale of MMC-Instruction (e.g., 1M vs. 600k instances) on the robustness of LMMs to distributional shifts in chart types, as measured by accuracy on unseen. 10 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MMC: Advancing Multimodal Chart Understanding with Large-scale Instruction Tuning. Research question: What is the impact of increasing the scale of MMC-Instruction (e.g., 1M vs. 600k instances) on the robustness of LMMs to distributional shifts in chart types, as measured by accuracy on unseen datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.7/10.

3 Results

13 papers retrieved. 10 claims extracted; 2 independently verified. Quality review score: 5.7/10.

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
MMC-Benchmark is a comprehensive human-annotated benchmark with nine distinct tasks evaluating reasoning capabilities over	✓	0.31
Extensive experiments on MMC-Benchmark reveal the limitations of existing LLMs on correctly interpreting charts, even for	✓	0.33
MMC-Instruction is larger and more diverse compared to existing chart question-answering datasets.	×	0.10
MMC-Instruction includes 600k free-form, open-ended questions with multiple-choice question answering (MQA).	×	0.06
MMCA achieves state-of-the-art performance on current chart question-answer benchmarks compared with existing open-source	×	0.11
MMC-Benchmark is the first human-annotated benchmark in line with human cognition to evaluate LLM’s ability to comprehend	×	0.11
MMC-Benchmark contains a wide range of tasks, including chart information extraction, chart reasoning, contextual chart	×	0.09
MMC-Benchmark offers two quantitative evaluation methods: free-format Generation Ability Evaluation using GPT-4 and multi	×	0.08
Experiments indicate that MMC-Benchmark poses significant challenges to GPT-4V, especially in Chart to Datable and Chart	×	0.10
MMCA model outperforms the model without fine-tuning the vision encoder in ChartQA, DocVQA, and TextVQA benchmarks.	×	0.05

References

- <http://arxiv.org/abs/2311.10774v2>
- <http://arxiv.org/abs/2504.20900v1>
- <http://arxiv.org/abs/2406.09961v2>