Interleaved Visual Reasoning Chains Enhance Transformer Performance in Multi-Step Logic Tasks

Assignee Research

June 5, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What architectural innovations improve transformer performance on multi-step logical reasoning v7. 20 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Zebra-CoT: A Dataset for Interleaved Vision Language Reasoning. Research question: What architectural innovations improve transformer performance on multi-step logical reasoning v7.

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 3.7/10.

3 Results

12 papers retrieved. 20 claims extracted; 1 independently verified. Quality review score: 3.7/10.

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
Anole with CoT prompting achieves 13.80% on MathVision.	×	0.04
Anole with CoT prompting achieves 22.80% on MathVista.	×	0.04
Anole with CoT prompting achieves 8.50% on VisuLogic.	×	0.05
Anole with CoT prompting achieves 12.80% on EMMA.	×	0.04
Anole with CoT prompting achieves 10.00% on MMVP.	×	0.04
Anole with CoT prompting achieves 26.46% on Blink.	×	0.04
Anole with CoT prompting achieves 23.60% on Vstar.	×	0.04
Anole–Zebra-CoT achieves 16.45% on MathVision.	×	0.05
Anole–Zebra-CoT achieves 25.30% on MathVista.	×	0.05
Anole–Zebra-CoT achieves 21.80% on VisuLogic.	×	0.05
Anole–Zebra-CoT achieves 15.02% on EMMA.	×	0.05
Anole–Zebra-CoT achieves 15.33% on MMVP.	×	0.05
Anole–Zebra-CoT achieves 31.25% on Blink.	×	0.05
Anole–Zebra-CoT achieves 27.20% on Vstar.	×	0.05
The base Anole model with chain-of-thought prompting is evaluated on mini versions of MathVision and MathVista.	×	0.05
Interleaved generation is time-consuming.	×	0.03
A full breakdown of each evaluation set is presented in Section C.	×	0.03
The example shows an interleaved visual reasoning chain generated by Bagel-Zebra-CoT.	✓	0.18
The first example involves subtracting all cylinders and adding 1 red sphere, resulting in 5 objects left.	×	0.01
The second example involves a 5x5 grid of small squares with 10 missing squares.	×	0.03

References

- http://arxiv.org/abs/2601.01982v1
- http://arxiv.org/abs/1609.04846v1
- http://arxiv.org/abs/2507.16746v2