

# Frontier Language Model Failures in Abstract Mathematical Reasoning

Assignee Research

June 6, 2026

## **Abstract**

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: What are the failure modes of frontier language models on abstract mathematical reasoning v12. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

## **1 Introduction**

This paper examines: ChaosBench-Logic: A Benchmark for Logical and Symbolic Reasoning on Chaotic Dynamical Systems. Research question: What are the failure modes of frontier language models on abstract mathematical reasoning v12.

## **2 Methodology**

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.0/10.

## **3 Results**

4 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.0/10.

## **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/2309.02144v1>
- <http://arxiv.org/abs/2601.01982v1>
- <http://arxiv.org/abs/2510.00071v2>