

Gemma-2 9B Benchmark Performance Across Reasoning Mathematics and Language Tasks

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: What are the benchmark performance scores of Gemma9B on reasoning mathematics coding and language understanding tasks. 7 claims were extracted from source literature; 7 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Crabs: Consuming Resource via Auto-generation for LLM-DoS Attack under Black-box Settings. Research question: What are the benchmark performance scores of Gemma9B on reasoning mathematics coding and language understanding tasks.

2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.3/10.

3 Results

3 papers retrieved. 7 claims extracted; 7 independently verified. Quality review score: 8.3/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
Existing studies on LLM-DoS attacks predominantly focus on white-box attacks, leaving black-box scenarios underexplored.	✓	0.39
AutoDoS is an automated algorithm designed for black-box LLMs that constructs a DoS Attack Tree and expands node coverage	✓	0.34
AutoDoS uses transferability-driven iterative optimization to work across different models using a single prompt.	✓	0.16
Embedding the Length Trojan allows AutoDoS to bypass existing defenses more effectively.	✓	0.25
Experimental results show that AutoDoS amplifies service response latency by over 250.	✓	0.21
AutoDoS leads to severe resource consumption in terms of GPU utilization and memory usage.	✓	0.21
The code for AutoDoS is available at https://github.com/shuita2333/AutoDoS .	✓	0.18

References

- <https://doi.org/10.18653/v1/2025.findings-acl.580>
- <https://doi.org/10.1007/s44443-025-00041-2>
- <https://doi.org/10.48550/arxiv.2412.13879>