

GPT2-Medium Benchmark Performance Across Reasoning Mathematics and Language Tasks

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What are the benchmark performance scores of GPT2-medium on reasoning mathematics coding and language understanding tasks. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.8/10. 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 benchmark performance scores of GPT2-medium on reasoning mathematics coding and language understanding tasks.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.8/10.

3 Results

14 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 2.8/10.

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
Deterministic systems can have sensitive dependence on initial conditions and still be labeled deterministic.	×	0.04
Chaotic systems are annotated with a positive largest Lyapunov exponent.	×	0.04
The system evolution is fully specified by its state and update rule in deterministic systems.	×	0.03
Random systems are defined by explicit noise terms or stochastic differential equations.	×	0.01
Chaotic systems are associated with canonical parameter regimes widely treated as chaotic in the literature.	×	0.03
PointUnpredictable captures practical loss of long-horizon point forecasts even under perfect modeling.	×	0.01
StatPredictable reflects that coarse statistical structure can remain predictable or stable over time.	×	0.02
Periodic, QuasiPeriodic, and FixedPointAttr are mutually exclusive indicators of common non-chaotic behavior classes.	×	0.01

References

- <http://arxiv.org/abs/2601.01982v1>
- <http://arxiv.org/abs/2410.12381v3>
- <http://arxiv.org/abs/2407.04973v1>