

Grok-4.1 Benchmark Performance Across Reasoning Mathematics Coding and Language Tasks

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: What are the benchmark performance scores of Grok-4.1 on reasoning mathematics coding and language understanding tasks. 7 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Granite Code Models: A Family of Open Foundation Models for Code Intelligence. Research question: What are the benchmark performance scores of Grok-4.1 on reasoning mathematics coding and language understanding tasks.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.4/10.

3 Results

8 papers retrieved. 7 claims extracted; 6 independently verified. Quality review score: 7.4/10.

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
The Granite series consists of decoder-only code models designed for code generative tasks.	✓	0.19
The Granite Code models were trained with code written in 116 programming languages.	✓	0.26
The Granite Code models family includes models ranging in size from 3 to 34 billion parameters.	✓	0.27
Evaluation on a comprehensive set of tasks demonstrates that Granite Code models consistently reach state-of-the-art per	✓	0.33
The Granite Code model family was optimized for enterprise software development workflows.	✓	0.32
All Granite Code models are released under an Apache 2.0 license.	×	0.15
The Apache 2.0 license for Granite Code models permits both research and commercial use.	✓	0.19

References

- <https://doi.org/10.3390/fi17090412>
- <https://doi.org/10.48550/arxiv.2405.04324>
- <https://doi.org/10.48550/arxiv.2301.05217>