

DeepSeek R1 Cross-Domain Code Generation Transferability on DS-1000 Across Languages and Complexities

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How does the cross-domain transferability of Deepseek R1's code generation performance compare to other LLMs when evaluated on the DS-1000 benchmark across programming languages with varying cyclomatic complexity distributions? 12 claims were extracted from source literature; 8 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: Neural Methods for Programming: A Comprehensive Survey and Future Directions. Research question: How does the cross-domain transferability of Deepseek R1's code generation performance compare to other LLMs when evaluated on the DS-1000 benchmark across programming languages with varying cyclomatic complexity distributions?.

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 7.4/10.

3 Results

4 papers retrieved. 12 claims extracted; 8 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 study synthesizes insights from more than 250 scientific papers.	✓	0.19
The majority of the papers surveyed were published between 2015 and 2025.	×	0.11
The survey includes foundational works dating back to the late 1990s.	×	0.14
The analysis spans 18 major programming tasks.	✓	0.21
The surveyed programming tasks include code generation, code translation, code clone detection, code classification, and	✓	0.27
The survey is guided by six research questions.	×	0.12
The survey traces progress from early code modeling approaches to advanced Code-specific Large Language Models (Code LLM	✓	0.30
The survey emphasizes the advantages of Code LLMs over traditional rule-based and statistical methods.	✓	0.19
The survey presents a taxonomy of evaluation metrics.	✓	0.16
The survey provides a categorized summary of datasets and benchmarks.	×	0.14
The survey identifies persistent limitations in data coverage and evaluation practices.	✓	0.18
The review distinguishes between neural models designed for natural language processing and those designed for programmi	✓	0.25

References

- <https://doi.org/10.3390/app152212150>
- <https://doi.org/10.48550/arxiv.2501.11223>

- <https://doi.org/10.48550/arxiv.2410.04466>