

Scalural Scaling Effects on Diagram-Based Coding Performance in aiXcoder Models

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does the scaling of model size (e.g., 27B vs. 7B) impact performance on HumanEval-V's diagram-based coding tasks, as evaluated by F1 score and task completion rate. 14 claims were extracted from source literature; 5 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: aiXcoder-7B: A Lightweight and Effective Large Language Model for Code Processing. Research question: How does the scaling of model size (e.g., 27B vs. 7B) impact performance on HumanEval-V's diagram-based coding tasks, as evaluated by F1 score and task completion rate?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.7/10.

3 Results

9 papers retrieved. 14 claims extracted; 5 independently verified. Quality review score: 6.7/10.

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
aiXcoder-7B generates code that is closer in length to human-written reference code compared to existing LLMs.	×	0.10
aiXcoder-7B is a lightweight and effective LLM with 7 billion parameters for code completion.	✓	0.27
aiXcoder-7B has received 2,193 GitHub Stars as of the submission date.	×	0.03
aiXcoder-7B proposes a novel training objective called Structured Fill-In-the-Middle, which considers the syntax structure.	✓	0.16
aiXcoder-7B proposes a new data sampling algorithm for code, which considers inter-file relationships and enhances the coverage.	✓	0.23
aiXcoder-7B releases a new code completion benchmark consisting of 16,136 samples and covering four languages.	×	0.08
aiXcoder-7B substantially outperforms 7 LLMs with similar sizes and even surpasses 4 larger LLMs (15B and 34B).	✓	0.22
aiXcoder-7B collects and cleans 2.8TB of natural language data and 3.5TB of source code data through its data collection.	×	0.11
The pre-training data of aiXcoder-7B consists of two parts: natural language data and source code data.	×	0.09
Natural language data for aiXcoder-7B is collected from two public datasets: WuDaoCorpora and RefineWeb.	×	0.07
aiXcoder-7B is evaluated on six code completion benchmarks.	✓	0.16
aiXcoder-7B outperforms DeepSeekCoder-7B, Qwen2.5-Coder-7B, CodeLlama-13B, StarCoder-15B, StarCoder2-15B, and CodeLlama-34B.	×	0.15
HumanEval consists of 164 Python programming problems.	×	0.04
MBPP consists of 974 Python programming problems.	×	0.00

References

- <http://arxiv.org/abs/2410.13187v3>
- <http://arxiv.org/abs/2511.12188v1>
- <http://arxiv.org/abs/2310.06825v1>