

# Performance Comparison of aiXcoder-7B and 13B Models on HumanEval Python Benchmark

Assignee Research

June 13, 2026

## Abstract

Large Language Models (LLMs) have been widely used in code completion, and researchers are focusing on scaling up LLMs to improve their accuracy. However, larger LLMs have lower inference efficiency, affecting developers' experience and productivity. In this paper, we propose a lightweight and effective LLM for code completion named aiXcoder-7B. Compared to existing LLMs, aiXcoder-7B achieves higher code completion accuracy while having smaller scales (i.e., 7 billion parameters). We attribute the superiority of aiXcoder-7B to three key factors: (1) Multi-objective training. We employ three tr

## 1 Introduction

This paper examines: aiXcoder-7B: A Lightweight and Effective Large Language Model for Code Processing. Research question: How does the pass@1 metric of aiXcoder-7B compare to 13B parameter models on the HumanEval Python benchmark?.

## 2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

## 3 Results

10 papers retrieved. 12 claims extracted; 9 independently verified. Quality review score: 7.5/10.

## 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.	✓	0.19
aiXcoder-7B is a lightweight and effective LLM with 7 billion parameters for code completion.	✓	0.22
aiXcoder-7B has received 2,193 GitHub Stars as of the submission date.	✓	0.18
aiXcoder-7B proposes a novel training objective called Structured Fill-In-the-Middle, which considers the syntax structure.	✓	0.19
aiXcoder-7B proposes a new data sampling algorithm for code, which considers inter-file relationships and enhances the code quality.	✓	0.28
aiXcoder-7B releases a new code completion benchmark consisting of 16,136 samples and covering four languages.	✓	0.22
aiXcoder-7B outperforms 7 LLMs with similar sizes and surpasses 4 larger LLMs (15B and 34B) in six code completion benchmarks.	✓	0.26
The pre-training data of aiXcoder-7B consists of 2.8TB of natural language data and 3.5TB of source code data.	✓	0.32
The natural language data for aiXcoder-7B is collected from WuDaoCorpora and RefineWeb datasets.	✓	0.19
aiXcoder-7B is evaluated on three popular NL2Code benchmarks: HumanEval and MBPP.	×	0.07
HumanEval consists of 164 Python programming problems.	×	0.04
MBPP consists of 974 Python programming problems.	×	0.04

## References

- <http://arxiv.org/abs/2508.15478v2>
- <http://arxiv.org/abs/2410.13187v3>
- <http://arxiv.org/abs/2410.12381v3>