# Impact of Multi-Source Knowledge Base Diversity on Reasoning in Sub-10B Parameter Retrieval-Augmented Code Generation

Assignee Research

June 8, 2026

**Abstract**

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: What is the effect of multi-source knowledge base diversity on the reasoning capabilities of sub-10B parameter models in retrieval-augmented code generation, as measured by pass@1 on HumanEval-Java. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: SIMCOPILOT: Evaluating Large Language Models for Copilot-Style Code Generation. Research question: What is the effect of multi-source knowledge base diversity on the reasoning capabilities of sub-10B parameter models in retrieval-augmented code generation, as measured by pass@1 on HumanEval-Java?.

## 2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

## 3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.7/10.

## 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.

## References


- http://arxiv.org/abs/2402.12317v2
- http://arxiv.org/abs/2505.21514v1
- http://arxiv.org/abs/2405.19265v1