Synchronous Query-Knowledge Base Evolution in EVOR for Semantic Preservation in Cross-Lingual Code Generation

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does synchronous query-knowledge base evolution in EVOR maintain semantic preservation compared to static RAG approaches on cross-lingual code generation tasks. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: EVOR: Evolving Retrieval for Code Generation. Research question: How does synchronous query-knowledge base evolution in EVOR maintain semantic preservation compared to static RAG approaches on cross-lingual code generation tasks.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.0/10.

3 Results

14 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 4.0/10.

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
With CodeLlama, the improvements of MPSC, ExeDec, and Reflexion are smaller than 2% on average compared to vanilla gener	×	0.08
The execution accuracy remains 0 in the Ring dataset across MPSC, ExeDec, and Reflexion methods when using CodeLlama.	×	0.04
Existing approaches excel in code generation tasks that do not require external knowledge, such as HumanEval.	×	0.12
DocPrompting significantly surpasses MPSC, ExeDec, and Reflexion by explicitly using documentation.	×	0.02
EVOR achieves a 16.1% absolute gain over DocPrompting when using ChatGPT.	×	0.06
EVOR achieves a 16.2% absolute gain over DocPrompting when using CodeLlama.	×	0.07
DocPrompting uses documentation as a single retrieval source without evolution in both queries and knowledge.	×	0.13
The paper uses execution accuracy (pass@1) as the default metric throughout the experiments.	×	0.04
MPSC incorporates both inter- and intra consistency by prompting LLMs to generate diverse outputs from Solution, Specifi	×	0.04
ExeDec employs a subgoal model to predict the subgoal of the desired program state and a synthesizer model to generate t	×	0.04
Retrieval-augmented code generation introduces risks of biased or incorrect information being retrieved, which could pro	×	0.09
There are privacy and security concerns if sensitive code snippets are inadvertently included in the retrieval process.	×	0.03

References

- http://arxiv.org/abs/2402.12317v2
- http://arxiv.org/abs/2508.11273v2
- http://arxiv.org/abs/2208.08227v4