

SWE-Shepherd Step-Level Feedback on Autonomous Coding Agent Efficiency

Assignee Research

June 3, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the impact of integrating fine-grained intermediate feedback from PRMs on the inference efficiency and token consumption of autonomous coding agents on the SWE-bench dataset. 13 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: SWE-Shepherd: Advancing PRMs for Reinforcing Code Agents. Research question: What is the impact of integrating fine-grained intermediate feedback from PRMs on the inference efficiency and token consumption of autonomous coding agents on the SWE-bench dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.5/10.

3 Results

11 papers retrieved. 13 claims extracted; 1 independently verified. Quality review score: 4.5/10.

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
SWE-Shepherd is evaluated on 100 tasks sampled from SWE-Bench Verified.	×	0.13
Each task requires generating a patch that resolves the issue and passes all associated tests.	×	0.03
The agent is limited to a maximum of 30 interaction steps.	×	0.03
SWE-Shepherd is compared against mini-SWE-Agent and SWE-Search.	×	0.05
mini-SWE-Agent is a strong LLM agent that follows a sequential decision process without explicit search or reward models	×	0.04
SWE-Search is a search-based framework that augments software agents with Monte Carlo Tree Search (MCTS) to explore multiple	×	0.06
SWE-Shepherd is mini-SWE-Agent augmented with the trained Process Reward Model (PRM) for action scoring and selection.	×	0.10
SWE-Shepherd converts agent trajectories into dense, step-level supervision by assigning scalar rewards to intermediate	✓	0.16
A dataset of action-level reward annotations is constructed using trajectories collected from SWE-Bench.	×	0.14
A lightweight reward model is trained on top of a base LLM using the constructed dataset.	×	0.12
At inference time, the PRM evaluates multiple candidate actions and guides the agent toward those predicted to be more useful	×	0.14
SWE-Shepherd enables reward-guided search without requiring full reinforcement learning.	×	0.11
SWE-Shepherd is positioned as a practical middle ground between supervised imitation learning and RL-based optimization.	×	0.04

References

- <http://arxiv.org/abs/2604.10493v1>
- <http://arxiv.org/abs/2411.07586v1>
- <http://arxiv.org/abs/2408.11296v1>