

V2.0 Black-Box Testing Report

AI Agent Functional Test

Tester: Swar Patel

Date: 2025-11-29

Version: 2.0

Test Repo Link: <https://github.com/Swar132/BlackboxTestingV2>

Summary of Findings

This test report validates the **stability and functional output** of the **AI Review Agent (v2.0)**. The system was subjected to a **manual load test of 15 unique Pull Requests**.

Metric	Result	Summary
System Stability	PASS ✓	Processed all 15 PRs without crashing
Functional Output	PASS ✓	Generated correct review reports with actionable feedback
AI Review Quality	PASS ✓	Consistently provided code examples and refactoring steps

The system is **STABLE** and **HIGHLY FUNCTIONAL**, consistently identifying valid **code bugs, security risks, and performance issues**.

Feature Status

Feature	Status	Notes
1. System Stability (Load Test)	PASS ✓	Successfully processed 15/15 PRs , all review reports generated
2. Functional Output (AI Review)	PASS ✓	Generated 15+ review files , bug detection accurate
3. Actionability	PASS ✓	Reviews consistently included refactored code examples

Detailed Test Cases

Test 1: System Stability & Load Testing

- Objective:** Validate batch processing reliability for diverse defect types

- **Action:** Executed review pipeline for PRs #1 – #15
- **Result:** **PASS ✓** – All `review_report_PR*.md` files generated successfully

Test 2: Functional Output & AI Quality

- **Objective:** Validate technical relevance and accuracy of code reviews
- **Action:** Spot-checked all generated reports
- **Result:** **PASS ✓** – Correctly identified all intended bug classes and provided specific fix suggestions

Key Observations from AI Reviews

PRs	Findings
1, 2, 9	Correctly identified critical runtime errors (<code>ZeroDivisionError</code> , <code>NameError</code> , <code>IndexError</code>)
11, 12, 13	Detected security risks & unsafe I/O (hardcoded API key, missing file checks, missing network error handling)
3, 4, 14	Flagged performance issues & side effects (<code>list mutation</code> , <code>time.sleep</code> , inefficient loops)
5, 8, 15	Found validation & typing issues (weak regex, missing type hints, ambiguous type comparison)
6, 7, 10	Detected code quality concerns (empty classes, redundant logic, magic numbers)

Bugs & Issues Found

Issue 1: Static Analysis Integration Failure (High Severity)

- **Description:** Across nearly all reviews (PR 1-15), the Static Analysis section reported: “issues with the file not being found” or “unable to detect the file”.
- **Impact:** Pylint, Flake8, and Bandit are failing to scan the generated files, likely due to path configuration in the test harness.
- **Suggested Mitigation:** Verify the `pr_files/` directory path passed to the subprocess calls and ensure files are flushed to disk before analysis runs.

Issue 2: Repetitive Newline Flags (Low Severity)

- **Description:** The AI flagged “Missing newline at end of file” for every single PR.
- **Impact:** Creates repetitive noise in the review output.
- **Suggested Mitigation:** Update the synthetic PR generator to automatically append a trailing newline (`\n`) to all generated files.

Recommendations

- **Fix Static Analysis Paths:** Debug the file path passing logic to ensure linters can see the temporary PR files.
- **Automate Formatting:** Fix the synthetic generator to comply with PEP 8 (newlines) automatically to reduce review noise.
- **Maintain Actionability:** The “Example Refactored Code” sections are excellent; ensure this structure is preserved in future prompts.

Conclusion

The **AI Review Agent v2.0 is stable, functional, and production-ready** regarding its core Logic Analysis capabilities. Reviews demonstrate strong bug identification performance. However, the **Static Analysis integration requires immediate debugging** to be effective.

Overall Status: READY (Pending Static Analysis Fix)
--