

Problem Statement

- Modern software development relies on continuous integration with thousands of automated tests.
- Tests cannot be run after each code change, typically running once daily.
- When tests fail, developers spend hours investigating root causes through logs and code changes.

Solution

The WhoBrokeMyTest AI Agent addresses the challenge of identifying the specific code changes responsible for test failures within the investigation process (Figure 1) by combining reasoning capabilities of Large Language Models (LLMs) with specialized tools (Figure 2).

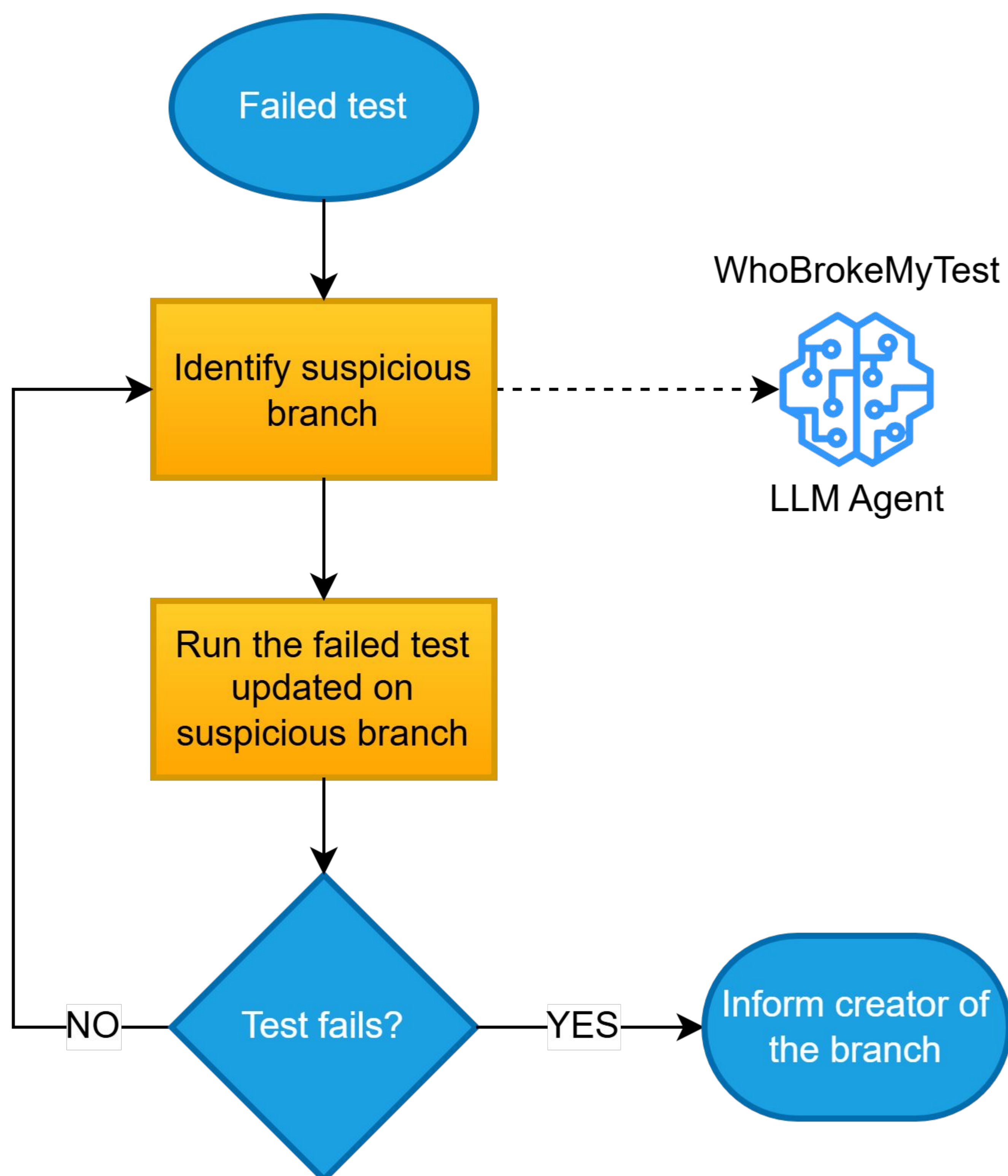


Figure 1: Failed test investigation process

WhoBrokeMyTest can handle queries such as:

- What branch was committed on the day of 13th of January that broke the X automated test?
- If I modify files X and Y, which automated tests do I need to run to make sure I don't break anything?

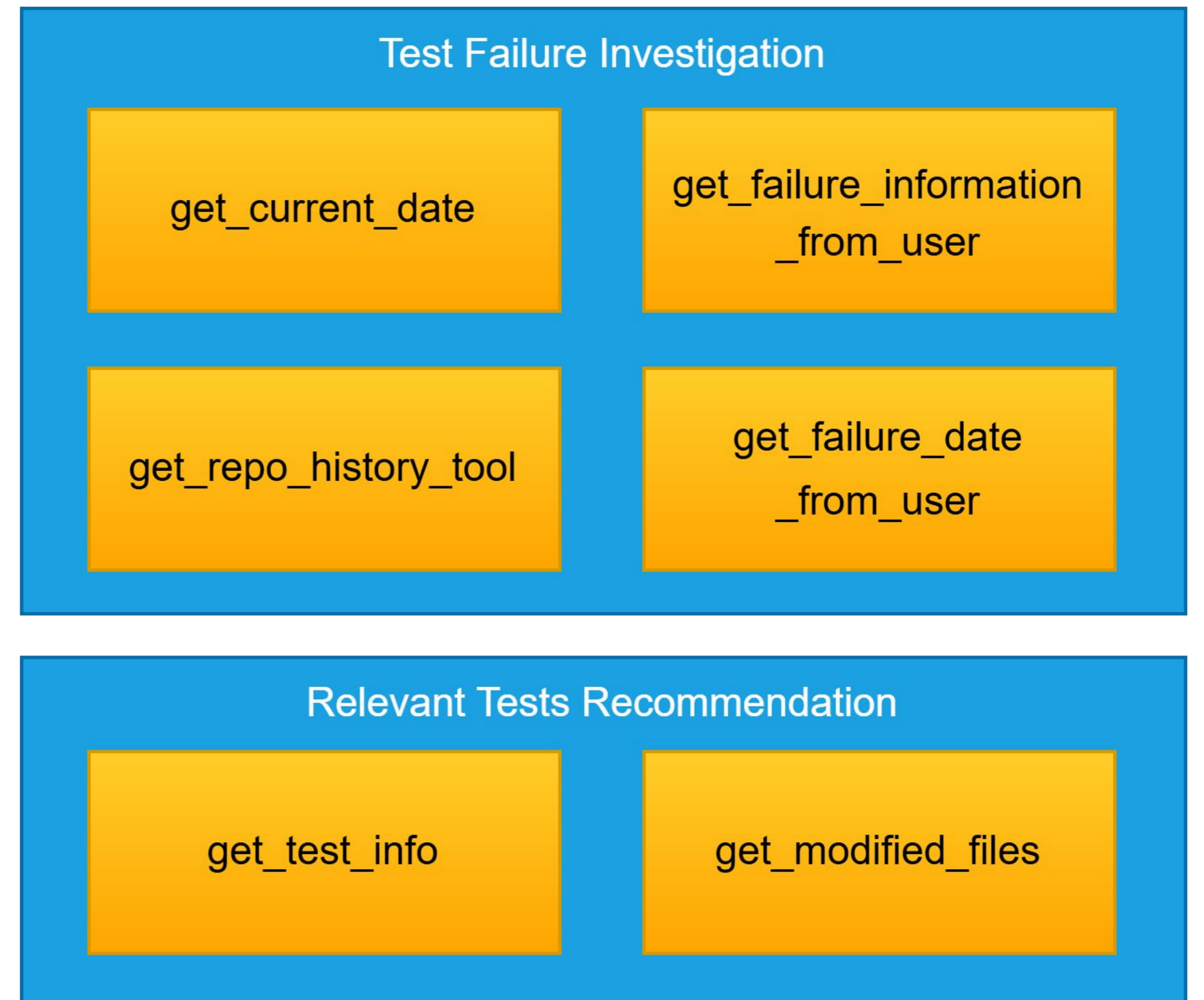


Figure 2: Tools used by the WhoBrokeMyTest agent

Productivity Impact

In complex multi-team environments, each development team can reclaim approximately 10 days per release cycle previously lost to test failure investigations.

Limitations

- Semantic matching constraints: Performance depends on textual similarities between errors and repository elements (commit messages, branch names, modified filenames).
- Test coverage gaps: Limited test documentation affects retrieval accuracy; enhanced coverage mapping would improve result relevance.

References

1. Robert Feldt, Sungmin Kang, Juyeon Yoon, and Shin Yoo. 2023. Towards Autonomous Testing Agents via Conversational Large Language Models. arXiv:2306.05152 [cs.SE] <https://arxiv.org/abs/2306.05152>
2. Mohamed Amine Ferrag, Norbert Tihanyi, and Merouane Debbah. 2025. From LLM Reasoning to Autonomous AI Agents: A Comprehensive Review. arXiv:2504.19678 [cs.AI] <https://arxiv.org/abs/2504.19678>
3. Betim Sherifi, Khaled Slhoub, and Fitzroy Nembhard. 2024. The Potential of LLMs in Automating Software Testing: From Generation to Reporting. arXiv:2501.00217 [cs.SE] <https://arxiv.org/abs/2501.00217>

