

PRESENTATION

Flakewatch: CI you can trust

We find, quarantine and explain flaky tests before they block another merge. Seed round, October 2026.



Made with MagicSlides

OVERVIEW

The problem

Flaky tests pass and fail on the same commit. Engineers learn to **hit re-run and ignore red builds**, and real regressions slip through with them.

Explore the **4 core pillars** that define this topic — each backed by research and real-world insight.

01

Wasted CI minutes

KEY

Re-runs triggered only because of flakes eat a large share of the monthly CI bill.

02

Blocked merges

CORE

A single flaky end-to-end test can hold a merge queue for an hour.

03

Lost trust

FOCUS

Once red builds are normal, teams stop reading failures at all.

04

No owner

IMPACT

Flakes live in shared suites, so nobody is on the hook to fix them.

What we measured in 12 repos

Numbers from 90 days of CI history across our 12 design partners.

[↑ Key Metrics](#)

7.8%

OF FAILED BUILDS WERE FLAKES

Same commit, same test, different result on re-run.

▲ Median across 12 repos

31 min

LOST PER ENGINEER PER WEEK

Waiting on re-runs and reading false failures.

▲ Self-reported, 140 engineers

64%

OF FLAKES WERE TIMING BUGS

Sleeps, race conditions and unawaited promises.

▲ Classified by our analyzer

-71%

RE-RUNS AFTER FLAKEWATCH

Drop in manual re-runs within 30 days of install.

▲ Average across partners

Flakes are **common, measurable and mostly fixable** once someone points at the right line.



Key Insight

Made with MagicSlides

How Flakewatch works

4-Step Framework

01

Ingest

Reads JUnit and CI results from every run.

- GitHub Actions
- GitLab CI
- Buildkite
- CircleCI

02

Detect

Scores each test on pass and fail history per commit.

- Same-commit retries
- Order dependence
- Runner and OS drift
- Time-of-day patterns

03

Quarantine

Flaky tests stop blocking merges but keep running.

- Auto skip-list PR
- Results still recorded
- Owner from CODEOWNERS
- Expiry after 14 days

04

Explain

Points to the likely cause with a suggested fix.

- Diff of passing vs failing logs
- Suspect line highlighted
- Fix PR from a coding agent
- Issue opened for owner

KEY INSIGHT

Install takes **one CI step and five minutes**, with no test code changes.

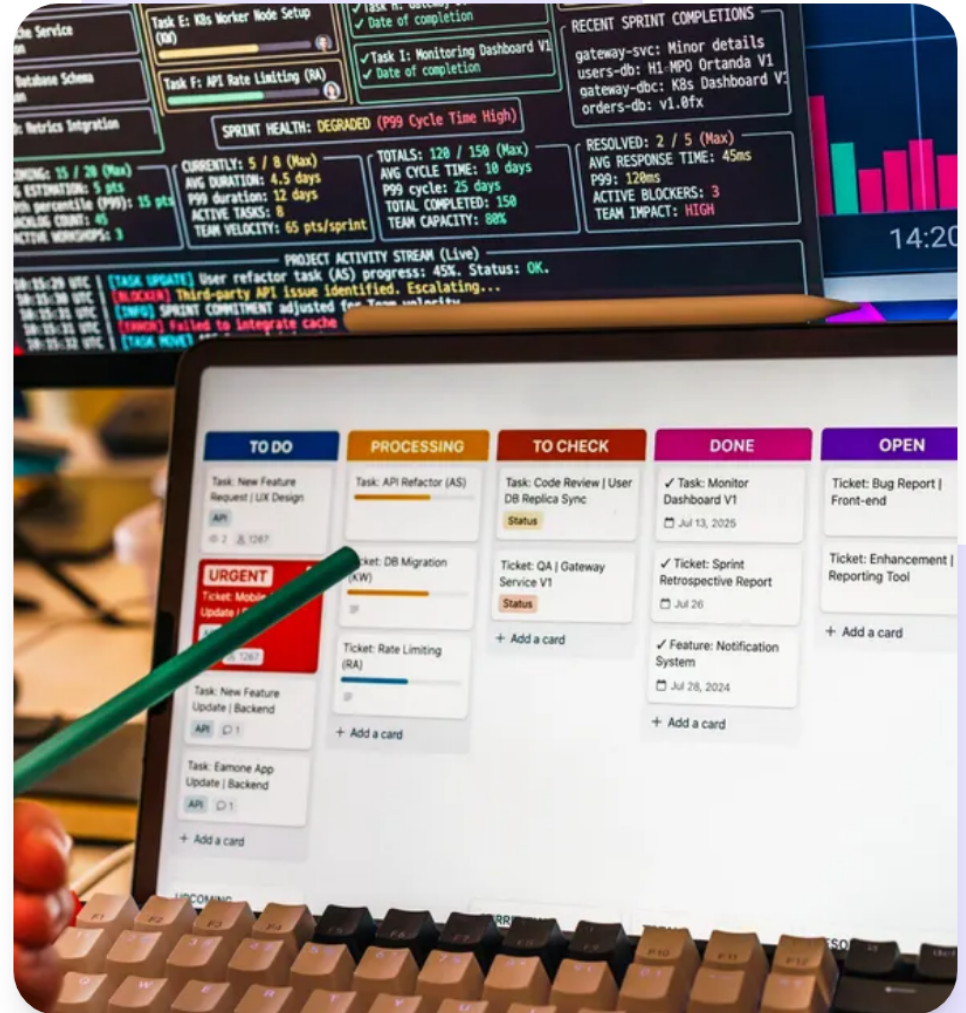


Made with MagicSlides

Traction and business model

Flakewatch is free for open-source repos and \$15 per committer per month for private ones. Nine of our 12 design partners converted to paid after the 30-day trial.

We are at \$11.4k MRR with 380 paying committers, growing 22% month on month. Our largest customer runs 9,000 CI jobs a day across 140 engineers.



THE ASK

\$2.5M seed

18 months to reach \$100k MRR and ship automatic fixes for the top flake types. Contact founders@flakewatch.example.

- Evidence-backed insights
- 4 proven advantages
- Ready to implement

1

Engineering, 55%

Four engineers on detection models and new CI integrations.

NEW

2

Go to market, 25%

Developer content, OSS sponsorships and one solutions engineer.

CORE

3

Infrastructure, 10%

Ingest pipeline for 50 million test results a day.

PRO

4

Reserve, 10%

SOC 2 Type II audit and runway buffer.

KEY



Made with MagicSlides