RUNSYBIL

MODEL · HARNESS · SYSTEM

What a pentest needs beyond the model

A better model makes the reasoning better. The harness lets it take actions. The system is what turns both into confirmed findings on your live application, and it is what Sybil is built to be.

What the test needs	MODEL **Frontier model alone**	MODEL + HARNESS **Model in a coding agent**	MODEL + HARNESS + SYSTEM **Sybil**
Reasons about code and flaws Reads code and responses, forms hypotheses	✓ Yes	✓ Yes	✓ **Yes**
Takes actions Runs tools, sends requests, decides what to try next	– No	✓ Yes	✓ **Yes**
Tests your live, running application Proves each issue against the real app, not just the source	– No	◑ Partly	✓ **Yes**
Keeps every request inside your scope Enforced by the system, not left to the model's judgment	– No	– No	✓ **Yes**
Tests as your real user roles Logs in and keeps sessions working for the whole engagement	– No	◑ Partly	✓ **Yes**
Independently reproduces each finding A separate check re-runs the exploit before you see it	– No	– No	✓ **Yes**
Deduplicates and reviews severity One finding per real issue, rated before it reaches you	– No	– No	✓ **Yes**
Remembers your application Each test starts from what earlier tests learned	– No	– No	✓ **Yes**
Retests fixes and tests continuously Confirms remediation and keeps pace as your code changes	– No	◑ Manual reruns	✓ **Yes**

Partly: possible with setup the customer builds and maintains. Describes general-purpose coding agents; specific tools and configurations vary.