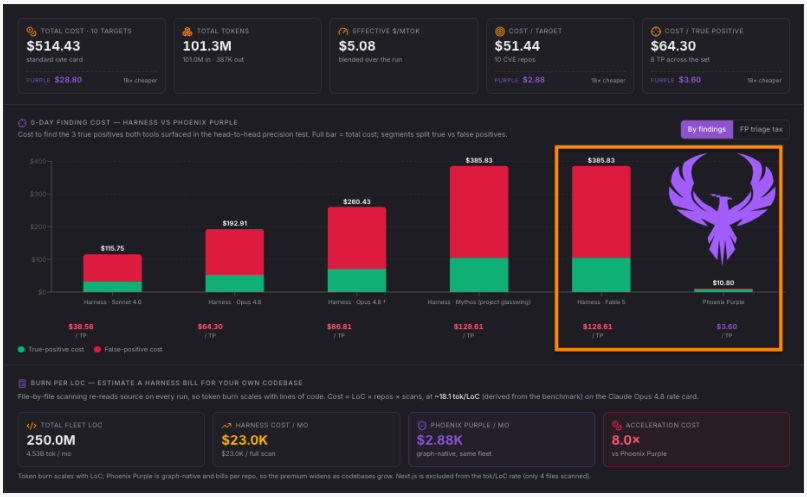
The Harness Is the Attack Surface - Phoenix Security launches Exploit Hunt at Black Hat USA 2026 to find zero day

Phoenix Security launches Exploit Hunt at Black Hat USA 2026: an AI red team that reports a finding only after it has written a working exploit for it



Phoenix Purple Token Saving Cost

LAS VEGAS, NV, UNITED STATES, August 6, 2026 /EINPresswire.com/ -- Phoenix Security today announced the general availability of Exploit Hunt, a capability within its Phoenix Purple platform that performs automated adversarial testing against application source code. Exploit Hunt reports a finding only after it has generated and validated a runnable proof-of-concept exploit for that finding. The capability is available to Phoenix Purple customers now and is being demonstrated at Black Hat USA 2026 this week.

> A scanner tells you a pattern looks risky. A pentester tells you what they had time to check, once a year. Exploit Hunt points a live threat model at the real attack surface without worry about hacks”
>
> *Francesco Cipollone*

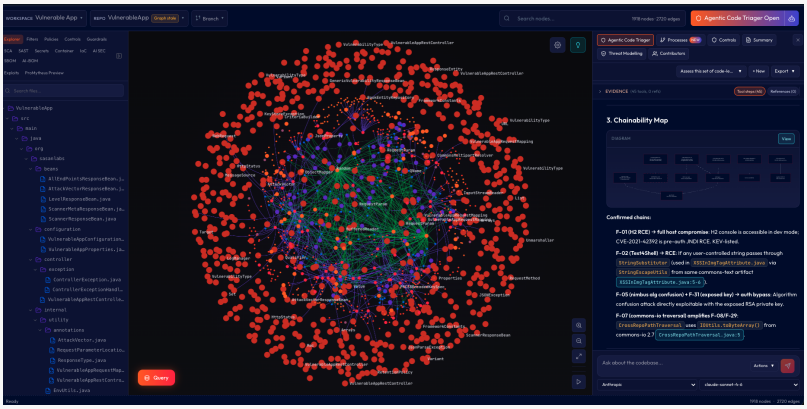
Exploit Hunt operates against a call graph, taint map, and knowledge graph that Phoenix builds for the repository before any model is invoked. It uses that structure to rank files for adversarial testing, weighing PageRank centrality, taint density, cyclomatic complexity, prior static analysis signal, external connectivity, and reachable dependency vulnerabilities. The workspace STRIDE threat model, including trust boundaries and attack-path chains, is supplied to the testing process as targeting input.

Three testing roles run in sequence, each with restricted visibility into the others' output. The first receives the source, the graph, and the threat model, and produces a vulnerability report containing a CWE classification, a severity rating, and an exploitation path. The second re-assesses that report independently, without the graph context that produced it, and returns a verdict of confirmed,

disputed, or insufficient evidence. The third attempt to produce a working exploit for candidates that survive, without access to the second role's verdict. Candidates who fail either stage are excluded from the report.

Each confirmed finding is delivered as a runnable exploit script and an accompanying validation document. Both are ingested into the platform's exploit store and are retrievable through the Phoenix Security API.

Exploit Hunt can be run on demand, as a native GitHub pull-request gate that posts a per-exploit comment and sets a commit status check, or on a daily, weekly, or monthly schedule. Exploit Hunt also makes the best use of hooks, enabling seamless integration with coding agents for a smooth developer experience: Developers can start a session, and Exploit Hunt indexes the current codebase and returns vulnerabilities and remedies at the end of the session. In this way, the agent can self-heal and remediate vulnerabilities before malicious code ever hits a merge.
When run on a schedule, files changed since the previous run receive a differential priority boost. Configured as a pull-request gate, a confirmed exploitable finding can block a merge.



Phoenix Security Purple Agentic SDLC

Phoenix Security Purple Attack Chain SCA agentic AI vulnerability

Research background

The methodology behind Exploit Hunt was developed during Phoenix Security research conducted in April 2026, following the accidental publication of a source map to the npm registry that exposed the reconstructed source of a widely used

developer command-line tool. Phoenix applied graph analysis to narrow an initial set of more than 100 static analysis findings to 8 candidates, then applied adversarial validation to those candidates.

Three vulnerabilities were confirmed, disclosed to the vendor, and subsequently registered as CVE-2026-35020, CVE-2026-35021, and CVE-2026-35022. All three are OS command injection issues classified under CWE-78. Phoenix Security published its full technical analysis on 6 April 2026. The complete write-up, including CVSS scores and affected versions, is available at phoenix.security/claude-code-leak-to-vulnerability-three-cves-in-claude-code-cli-and-the-chain-that-connects-them.

Gartner ASM ASPM Voice of the customer Phoenix Security

Scope

Exploit Hunt is not a replacement for human penetration testing but augments engineers' ability to find vulnerabilities, grouping them into low-friction changes and higher-friction, breaking changes that require more attention.

Operational controls

Exploit Hunt ships with a set of operational controls. Model API keys are bring-your-own by default. Each run is subject to a spend cap and is backed by a durable skip ledger. Live scanning requires explicit opt-in. Findings are grouped and deduplicated. Runs are organization-scoped and fail closed on foreign or malformed references. Remediation output is delivered as reviewable pull requests with supporting reasoning, and changes identified as potentially breaking require human approval before they are applied.

Advanced Capabilities

Exploit hunt also leverages existing SAST, SCA, and AI scans to determine whether they can be used to derive an attack, mixing and merging deterministic scanning capabilities with non-deterministic ones. Phoenix hunt also supports the early-release threat modeling lead hunt, mapping the attack surface and attack vector paths to inform the vulnerability chaining engine of

potential high-risk paths.
Exploit hunt can also ingest or generate architectural paths and patterns to detect whether an attacker could leverage architectural gaps, such as broken identity, to attack the application.
Phoenix Security Hunt memory is also available in early beta to remember previous attacks and previously determined false positives, enabling the following hunts to improve and learn from the past

Cost modelling

Phoenix Security publishes a public cost calculator for AI-assisted scanning at ai-scan-cost.phoenix.security. Using the assumptions published there, graph-native analysis models at approximately $3.60 per verified true positive, compared with approximately $64 per verified true positive for a file-by-file scanning approach applied to the same repositories using the same model. Modeled results vary with fleet size, code volume, and scan frequency. Figures are produced by the published calculator and are not measured customer results.

Executive comment

"We see Hunt as a leap forward focused on engineering instead of just providing them vulnerabilities. The low friction, high friction changes is a welcoming gift to engineering, enabling them to quickly determine if this (the change) needs attention or not"

"We built Exploit Hunt because the gap between a finding and a fix is usually the time spent proving the finding is real," said Francesco Cipollone, CEO & Co-Founder of Phoenix Security. "Our own research in April convinced us this could be systematized. We narrowed a hundred static findings down to eight, validated three of them, and the work that mattered was the validation rather than the detection. Exploit Hunt puts that sequence into the platform, so what reaches an engineer is an attack path with an attached script, rather than a severity score to investigate. Teams review and approve every change. Nothing is applied automatically."

Availability

Exploit Hunt is generally available in Phoenix Purple, including the native GitHub pull-request gate and scheduled runs. Product details are available at phoenix.security/phoenix-purple-ai-sast-sca-ai-generated-code. The Agentic SDLC Security control framework is documented at phoenix.security/agentic-sdlc-security-control-framework-three-pillars.

Phoenix Security will exhibit at Black Hat USA 2026 at Booth 5702, Mandalay Bay Convention Center, Las Vegas, in the Business Hall, from 4 to 6 August 2026. Demonstrations can be booked on the stand or at phoenix.security/request-a-demo.

Phil Moroni
Phoenix Security

+1 919-594-8888
email us here
Visit us on social media:
LinkedIn
Instagram
Facebook
YouTube
X
