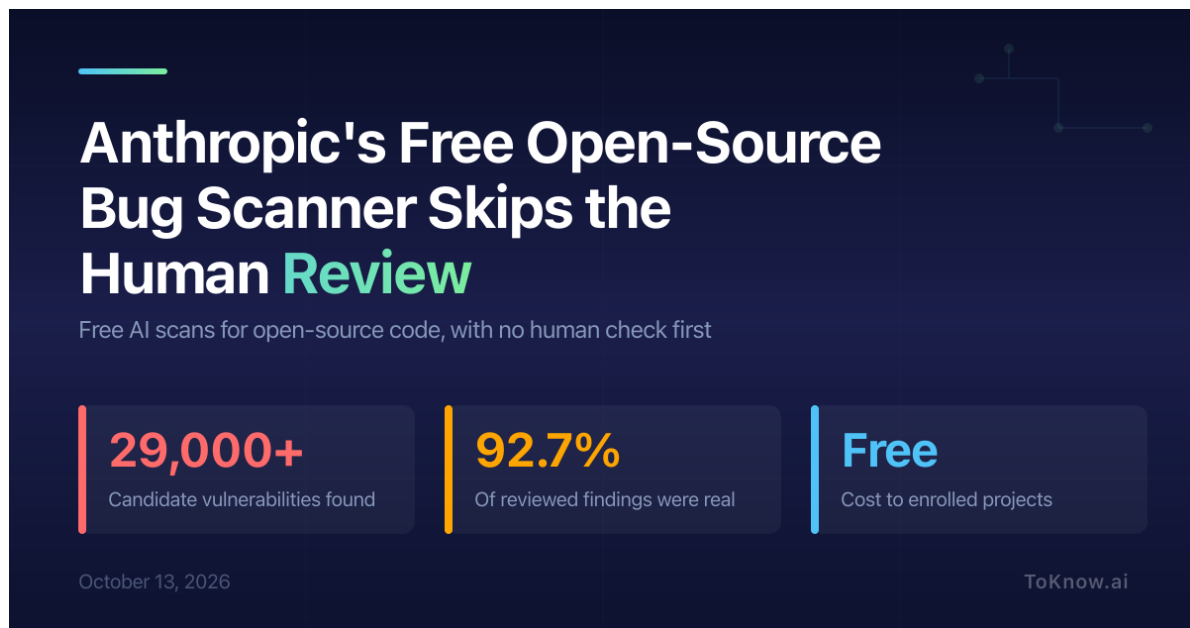


Anthropic's Free Open-Source Bug Scanner Skips the Human Review

Kabui, Charles

2026-10-13

[Read at ToKnow.ai](#)



The screenshot shows a dark-themed header for a blog post. At the top left, there is a small horizontal bar with a green-to-blue gradient. The main title 'Anthropic's Free Open-Source Bug Scanner Skips the Human Review' is in large white font, with 'Human Review' in green. Below the title is a subtitle 'Free AI scans for open-source code, with no human check first'. Three statistics are displayed in colored boxes: '29,000+' (red) for 'Candidate vulnerabilities found', '92.7%' (orange) for 'Of reviewed findings were real', and 'Free' (blue) for 'Cost to enrolled projects'. The date 'October 13, 2026' is at the bottom left, and 'ToKnow.ai' is at the bottom right. A faint network diagram is visible in the top right corner.

Anthropic's Free Open-Source Bug Scanner Skips the Human Review

Free AI scans for open-source code, with no human check first

29,000+ Candidate vulnerabilities found	92.7% Of reviewed findings were real	Free Cost to enrolled projects
---	--	--

October 13, 2026 ToKnow.ai

OSS Scanner is a free service from Anthropic that hunts for security bugs in open-source software and emails what it finds to the people who maintain that code. Maintainers opt in by opening a pull request on GitHub. Each report includes a working proof of concept, an explanation of the flaw, and a suggested fix. Every report is written by AI models, with no

human checking it first. Anthropic built it after a similar scan during Project Glasswing, and says Google's OSS-Fuzz inspired the design.

The reason for skipping the human check is volume. Anthropic's models found over 29,000 candidate vulnerabilities in widely used software over six months, and its team could only manually review about 6,000. On [CyberGym](#), a public test, the models went from finding under 20% of the vulnerabilities at the start of last year to over 85% this year. Finding bugs is now easier than checking them. Of the findings an outside security firm reviewed, 92.7% turned out to be real. The software library wolfSSL said 72 of the 74 reports it received were valid, and five became official security advisories.

Some of the unreviewed reports will be wrong, like an inflated severity rating, so Anthropic is not imposing a 90-day disclosure deadline by default. Anthropic frames the push as a race, since capable models are [now widely available to attackers too](#). The same launch created a program to protect power grids, water systems and transport, with 11 partners including CrowdStrike, Palo Alto Networks and Rockwell Automation.

Read More: [OpenAI Daybreak vs. Anthropic Mythos: two rival visions for AI cybersecurity](#)

Sources:

- [Introducing the Anthropic Cyber Mission \(Anthropic\)](#)
- [Launching an opt-in vulnerability-finding service for open-source software \(Anthropic Frontier Red Team\)](#)
- [OSS Scanner \(Anthropic\)](#)
- [Coordinated vulnerability disclosure dashboard \(Anthropic\)](#)
- [Anthropic Launches Free AI Vulnerability Scanner for Open-Source Projects \(The Hacker News\)](#)

Disclaimer: For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation. **Read more:** [/terms-of-service](#)