

## Hey! What's New? 2025-86

### AI Beats Hackers to a Zero-Day Cybersecurity Discovery, Twice

An article in *TechRepublic* says that AI prevented real-world cyberattacks before they began. “Google’s AI agent Big Sleep identified the critical vulnerability CVE-2025-6965 before cybercriminals could exploit it in the wild. And Microsoft’s Security Copilot uncovered a wave of bootloader flaws that could have allowed attackers to bypass Secure Boot protections across Linux systems. These instances mark a turning point: AI is now fast and capable enough to beat human threat actors to zero-day vulnerabilities.”

The article points out that “developed by Google DeepMind and Project Zero, Big Sleep identified a memory corruption issue in SQLite that affects all versions prior to 3.50.2. The vulnerability, rated 7.2 on the CVSS scale, allows attackers to exploit integer overflows and potentially read beyond array boundaries through crafted SQL inputs. Google’s Threat Intelligence team had already detected signs that hackers were staging a zero-day exploit but had not pinpointed the bug itself. Big Sleep did.”

“We believe this is the first time an AI agent has been used to directly foil efforts to exploit a vulnerability in the wild,” said Kent Walker, president of Global Affairs at Google and Alphabet.

According to the article, “SQLite maintainers confirmed the vulnerability was a serious issue known only to attackers before it was disclosed and patched. It may have been hidden in the codebase for years — undetectable by traditional fuzzing methods.”

It then points out that “Microsoft’s Security Copilot audited open-source bootloader code and found 11 vulnerabilities in GRUB2, the Linux bootloader used in many operating systems. Successful exploitation could bypass Secure Boot and allow persistent bootkit installation.” The AI also flagged several vulnerable functions related to filesystem mounting and accelerated vulnerability discovery in U-Boot (four flaws) and Barebox (five flaws). One of the most critical GRUB2 issues received a CVSS score of 7.8. All of the vulnerabilities were fixed by February 2025, but the speed and accuracy of discovery signal a new role for AI in securing foundational system software.”

Google’s internal OSS-Fuzz system, now enhanced with AI, found 26 new vulnerabilities and expanded test coverage across 160 projects by up to 29%. “One project saw a 7,000% increase in coverage, jumping from 77 lines to more than 5,400. Many of these bugs were found in codebases that had already undergone extensive fuzzing and testing over many years.”

The article also notes that Google reported significant real-world impact in 2024, suspending 39.2 million advertiser accounts using AI — triple the previous year. Deepfake ad reports dropped 90% thanks to large language model-powered detection systems.

“Security researchers noted that traditional fuzzing tools failed to detect the SQLite flaw that Big Sleep uncovered. Despite two decades of testing, the vulnerability had remained hidden. The difference lies in how AI agents interpret code. Instead of brute-forcing test inputs, models like Big Sleep recognize subtle patterns and contextual relationships that legacy tools miss.”

The scale advantage is becoming clear, the article stresses. “Ponemon Institute’s 2024 research shows organizations face more than 22,000 security alerts per week; AI can handle

over half of them without human input, yet more than 12,000 unknown threats still go undetected using conventional tools.”

Google is already adapting to this shift, it adds. “Its vulnerability rewards program now includes AI-specific attack categories like prompt injection and training data exfiltration. In the program’s first year, Google paid over \$50,000 for GenAI-related bugs. Google’s Bug Hunters team noted that approximately one in six reports resulted in actual product changes.”

Enterprise adoption is accelerating as well. The article says that “around 66% of organizations believe AI will improve security team productivity and 70% say it is already detecting threats that previously went unnoticed. Still, only 18% have fully deployed AI-based security tools, suggesting major growth ahead.”

Google reported in November 2024 that its updated OSS-Fuzz now covers 272 C/C++ projects, adding more than 370,000 lines of new test coverage and uncovering vulnerabilities that had slipped through traditional scanners. “These developments point to a larger transformation already underway. Big Sleep and Security Copilot demonstrate that zero-day discovery is shifting from a reactive process to a predictive one. Security teams can now scale their impact using AI agents, reduce time-to-discovery from months to hours and audit massive codebases more thoroughly than ever before.”

Learn more at [AI Beats Hackers to a Zero-Day Cybersecurity Discovery, Twice](#).