

AI-Accelerated Vulnerability Discovery

On June 22, 2026, OpenAI released the full version of GPT-5.5-Cyber, a specialized AI model designed to autonomously find software vulnerabilities, trace attack paths through large codebases, validate whether flaws are exploitable, and generate working patches. During pre-release testing, OpenAI researchers used the model to discover five exploitable vulnerabilities in Google Chrome, more than ten in Apple Safari, and a flaw in Firefox that Mozilla patched just two days before the Pwn2Own hacking competition, causing five of the six Firefox entries to withdraw.

OpenAI states that GPT-5.5-Cyber is restricted to vetted cybersecurity defenders through its Trusted Access program. However, the UK AI Safety Institute evaluated the standard GPT-5.5 model and reported that expert red teamers discovered a universal jailbreak that bypassed all of the model's safety restrictions in approximately six hours. The Canadian Centre for Cyber Security has separately warned that organizations should assume AI-driven exploitation will bypass preventative controls and outpace vendors' ability to publish fixes. As with Anthropic's Mythos release, the safeguards on these models are not keeping pace with the capabilities behind them.

This is the second frontier AI model in as many months to demonstrate the ability to find and exploit software vulnerabilities at machine speed. With attackers now able to find and exploit flaws faster than vendors can patch them, fundamentals like prompt patching, strong access controls, modern endpoint protection, and regular vulnerability scanning matter more than ever.

If you would like to discuss how this affects your organization, please respond to this email to reach me directly. We will review your environment and advise you on next steps.

Further information is available via the links provided below.

The Hacker News: OpenAI Expands Daybreak With GPT-5.5-Cyber

Axios: OpenAI Gives GPT-5.5-Cyber More Powerful Cybersecurity Capabilities

UK AI Safety Institute: Our Evaluation of OpenAI's GPT-5.5 Cyber Capabilities

Cyber Security News: OpenAI Releases GPT-5.5-Cyber With Full Automation for Vulnerability Detection

Ref. GPT-5.5-Cyber AI Capability Advisory, June 2026

Sincerely,

Chris Harper

Chief Security Officer

Sandbox Technologies, Inc.