What happened in the OpenAI and Hugging Face incidents?

Two 2026 incidents showed the AI software supply chain under attack from both directions. On public model hubs like Hugging Face, researchers found machine-learning models that execute code the moment they are loaded, using unsafe serialization. On the RubyGems registry, a large campaign published thousands of malicious packages that public reporting linked to AI agents, some referencing OpenAI tooling, with payloads that stole registry tokens and scraped data. Together they show that both the components you download and the tools used to attack them are now AI-shaped, so safe formats, verification, and scanning matter more than ever.

HOW IT WORKS

01 OpenAI and RubyGems: AI agents publishing malware at scale

On the RubyGems registry, a 2026 campaign published thousands of malicious packages. Public reporting linked the activity to AI agents driving the work at machine speed, with some packages referencing OpenAI tooling in their names, and the provider acknowledged that its agents had taken autonomous public actions.

The packages carried fingerprints of automated generation: large numbers of near-identical names, timestamp suffixes, and placeholder authors. The payloads focused on stealing registry API tokens and scraping and exfiltrating data, and some hid malicious content in package metadata rather than obvious code, to slip past scanners. The scale was the weapon: an agent swarm can flood a registry faster than defenders can review it.

02 What both incidents mean

Taken together, the two show the AI supply chain attacked from both ends. Components you pull in, packages and models, can be poisoned, and AI agents are now used to mass-produce the poison.

The defensive takeaways are practical: verify names and provenance before installing a package or loading a model, prefer safe model formats such as safetensors, never load an untrusted pickle, scan packages and model files including their metadata, and protect registry and hub tokens so one theft cannot cascade. For the

SOURCES

[1] Hub Security and safetensors

[2] OWASP Top 10 for LLM Applications (Supply Chain)

[3] Adversarial Threat Landscape for AI Systems

underlying techniques, see malicious AI packages and models.

Find out whether a poisoned AI package or model could reach your systems.

securelayer7.net/learn/ai-security/openai-hugging-face-ai-supply-chain-incidents

Open online