AI attacks are getting faster and more automated

Hackers have used AI for some time. What is new is how much of the attack AI now handles. ARTEX is the example.

What ARTEX is

An open-source security-testing tool that attackers have taken up. A person types a goal in plain words, and AI agents carry out the attack.

The FBI says

In some intrusions, ARTEX takes attackers from first target to privileged access in under three hours.

What it has done, and what it takes

- **Real damage.** Data on about 67,000 people was taken from South Korean financial institutions. A China-aligned state group, per Microsoft, used ARTEX against more than 2,250 ConnectWise downstream tenants.
- **Little skill needed.** People with minimal technical expertise can run complex attacks on many targets at once.
- **Many AI models plug in.** Claude, GPT and Chinese models all connect. Chinese ones are most common among attackers, helped by low cost and few guardrails.

Write-ups on ARTEX since 30 October (selected)

30 Oct Google: ARTEX and Friends
6 Nov Microsoft: ARTEX abuse expands to the financial sector
11 Nov Palo Alto Networks Unit 42: ARTEX and AI-powered red teaming tools
13 Nov FBI: flash alert on ARTEX and Cisco Talos: ARTEX and beyond

SOURCES FBI flash alert FLASH-20261113-001, 13 Nov 2026. Google Threat Intelligence Group, AI Threat Tracker, 16 Nov 2026.

Google puts attackers at stage 5 of 7 on the way to full automation

Google's own scale of how attackers use AI, 16 November. Stages 1 to 5 have been seen. It projects stages 6 and 7 within 1 to 6 months.

1	**Prompt-based use of LLMs** End of 2025 to early 2026	**Human-driven (chatbot)**
2	**Embedding LLM capabilities into malware** Early 2026	**Automated runtime call**
3	**Incorporating LLM-enabled automation into operations** Early 2026 to Sep 2026	**LLM-in-the-loop**
4	**Developing and experimenting with autonomous frameworks** Mid 2026 to Sep 2026	**Experimental autonomy**
5	**Multiple attack stages conducted by LLM-driven tooling in the wild** NOW Current stage	**Minimal human-in-the-loop**
6	**Multi-agent, unsophisticated attacks with little to no human interaction** Near term: 1 to 3 months	**No human-in-the-loop**
7	**Sophisticated, fully autonomous attacks** Medium term: 3 to 6 months	**Fully autonomous**

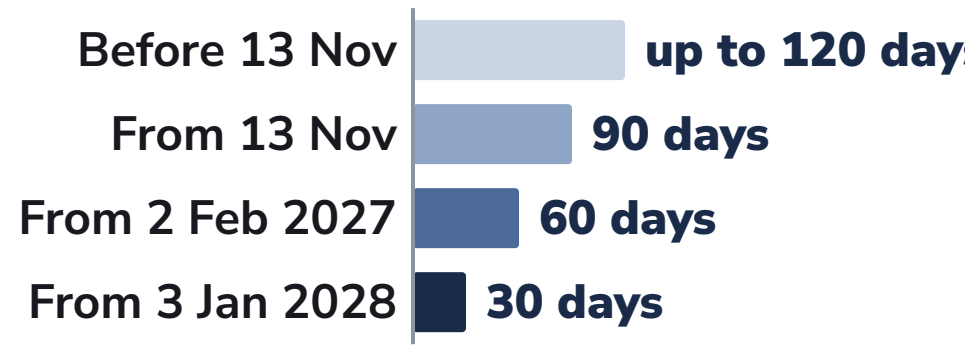
Seen in the wild so far: ARTEX (FBI and Google), and CAIRN and PentAGI (Google). Solid boxes are observed, dashed boxes are Google's projection.

SOURCE Google Threat Intelligence Group, AI Threat Tracker, 16 Nov 2026, Figure 2. Redrawn here with Google's wording shortened in places.

The time to fix a flaw is running out

Project Zero shortens its deadline

Project Zero is Google's security research team. It finds software flaws, reports them to the maker and publishes them after a deadline. Days to that deadline:



Until now: 90 days to fix, plus 30 for users to update.

- **Patch to attack, almost at once.** Google's bug team calls zero patch-to-exploit time “a reasonable, though imperfect, model” of attackers.
- **More flaws exploited.** Disclosures and exploits nearly doubled from January to August 2026. Zero-days rose only marginally, from 8 a month in 2025 to 11 in 2026, before a spike to 22 in August.
- **No human in the loop.** For the first time, Google found a tool that compares code before and after a patch, drafts an exploit and runs it.
- **AI helps defenders too.** Project Zero now finds several flaws a week where it once found a few a year.

UK cyber agency (NCSC): AI-powered malware is “highly likely” in a ransomware attack on UK critical infrastructure by 2028.

What has not changed

Google: the mechanics of intrusions “remain the same, such as using compromised credentials and targeting unpatched systems”.

Not fully automated yet

Google says these tools “currently trade stealth for velocity”. WIRED: fully AI-powered ransomware attacks have “yet to emerge”.

Earlier notes (tigzig.com ➜ Analysis)

21 Sep Sophisticated attacks no longer require sophisticated attackers

10 Sep Four warnings about AI and cyber risk arrived inside three weeks

This is my analysis of public sources, for information only. Care has been taken, but errors are possible. Use of AI. Parts of this work were prepared with AI under my direction. I reviewed, edited and approved every page. The analysis and conclusions are mine.

SOURCES Google Project Zero, 13 Nov. Google Threat Intelligence Group, 16 Nov. UK NCSC, 3 Nov. WIRED, 14 Nov 2026.