Defending against Artex: Ten steps against agentic AI attacks

Artex is an open-source agentic AI penetration-testing framework. In recent S-RM incidents, activity linked to Artex moved from a flaw in a web server to credentials, cloud and identity systems within hours, and came back within minutes of being contained. It worked on several systems at once, so defenders were responding to one compromise while the agents were already active in another. The weaknesses it used are old ones; what has changed is the speed at which they can be exploited. These ten steps are drawn from what we saw.

Artex attack speed

21 min
re-entry after the first server was contained

10 min
to move from a blocked IP to 50+ new addresses

~4 h
from first injection callback to operating-system commands

Ten key steps

1 KEEP IT OUT
Make the WAF the only front door
Accept traffic on load balancers and procies only from the WAF or CDN (private ALB, origin-restricted security group or secret header). Run the WAF in Block mode with header and body inspection.
Why: Artex-linked probing reached web tiers through load balancers, and header and body SQL injection passed WAF rules.

2 KEEP IT OUT
Validate file uploads by content
Check content, not just MIME type or extension. Store uploads where scripts cannot run, and review every upload handler, including legacy ones.
Why: attackers used double extensions and image/script polyglots to place web shells.

3 MAKE INJECTION HARMLESS
Disable `xp_cmdshell` on SQL servers
Turn off `xp_cmdshell`, `sp_OACreate` and `CLR`. Deny `xp_dirtree` to public and disable `sa`.
Why: SQL injection became operating-system commands about four hours after the first callback.

4 MAKE INJECTION HARMLESS
Restrict SQL and SQL Agent privileges
Give each site its own non-admin login. Run SQL and Agent as low-privilege accounts, restrict CmdExec job steps, and stop `sqlservr` and `sqlagent` launching `cmd` or PowerShell.
Why: Agent jobs and hidden database admin accounts were used to stay in.

5 SEE IT EARLY
Detect Artex-style names
Alert when files, accounts, jobs, VMs, keys or branches appear as `i<3-5 digits>` or `artex` / `artx` plus digits. Hunt back through existing logs too.
Why: Artex tags its work with rising counters. Names are easy to change, so pair this with behaviour: rapid key testing and parallel logins.

6 SEE IT EARLY
Detect and block `*.oast` callbacks
Alert on DNS or SMB from web and database servers to OAST domains such as `oast.fun`. Block outbound DNS and SMB from those tiers.
Why: out-of-band injection tests called out hours before commands ran, a window to act in.

7 STOP FOOTHOLDS STICKING
Enable web shell detection and blocking
Turn on AV and EDR web-shell modules and scan web and upload directories. Check the sensor is connected on every instance and on the base image.
Why: in two cases the hosts the attacker used most had no working EDR telemetry.

8 STOP FOOTHOLDS STICKING
Restrict lateral movement from web servers
Segment web servers from each other, from management systems and from the database tier (allow port 1433 only from web servers). Block outbound SMB and WinRM.
Why: the agents moved across sibling hosts and on to domain controllers within hours.

COPYRIGHT © 2026 S-RM

9 LIMIT THE SPREAD

Run secret scanners on web servers

Scan web roots, config files, shell history, build-system state files and disk images for hardcoded credentials. Move secrets to a vault and rotate anything found.

Why: credentials left in files were tested everywhere within minutes.

10 LIMIT THE SPREAD

Lock down cloud credentials on the web tier

Require IMDSv2, use least-privilege instance roles, and autorevoke on GuardDuty `InstanceCredentialExfiltration.OutsideAWS` alerts.

Why: instance-role keys were taken from a web server and used from the internet.

If you suspect Artex: first-hour actions and what to hunt for

Contain it

1. Contain the whole group, not the first host

Block the vulnerable path at the load balancer or WAF, then isolate every host built from the same image. Sibling servers running the same code let the attacker back in. Seen: re-entry within 21 minutes.

2. Remove access before you rotate

Find and delete the accounts, keys, jobs, VMs and deploy keys the attacker created. Rotating first hands over the new secrets. Make replacement keys less privileged. Seen: a rogue owner account survived a bulk identity reset. 3. Do not rely on IP blocks Addresses rotated between hosting ranges, VPNs and Tor. Use behaviour-based rules and rate limits. Blocking alone will not stop the activity. Seen: 50+ addresses within 10 minutes of a block.

3. Do not rely on IP blocks

Addresses rotated between hosting ranges, VPNs and Tor. Use behaviour-based rules and rate limits. Blocking alone will not stop the activity. Seen: 50+ addresses within 10 minutes of a block.

4. Preserve before you change

Stop auto-scaling from terminating instances, then snapshot disks and capture memory and logs. Turn on logging that was missing: load balancer, WAF, object-level storage, database audit. Seen: "no sign of download" is weak when logging was off.

Hunt for it

Where	What to look for
Counters	Scripts, VMs, keys, accounts and files named `i<3–5 digits>` , with numbers rising during the intrusion
Names	`artex` / `artx` plus digits on accounts, jobs, uploads and probe files; GitHub branches `artex-wtest-<n>` with a readme containing `probe`
Task folders	`/opt/artex/data/tasks/<n>/i<id>/…` on compromised hosts
Web requests	User agent `norma/0.4` among generic `curl` and Python agents; SQL strings starting `ARTEX'` ; placeholder emails `<id>@example.invalid`
Callbacks	DNS or SMB lookups from database and web servers to `*.oast` domains
Behaviour	Logins lasting one to two minutes with no shell history; `bash -lc` and sudo `bash -s` ; keys tested against many accounts within minutes; two or three attack chains in parallel; fast retries from sibling hosts after a block

Names and user agents are easy to change, and Artex is public, so a copycat could produce the same markers. A marker plus the behaviours above is a strong lead; a marker alone is weak. Absence of markers does not clear an environment.

Suspected cyber incident? Talk to an expert.

S-RM's Cyber Incident Response team works these cases. If you are concerned about your ability to detect or contain AI-driven intrusion activity, or want to test your response at this speed, we can help.

hello@s-rminform.com
www.s-rminform.com

Read the full article:
Artex in the wild: lessons from responding to agentic AI attacks

S-RM is a global intelligence and cyber security consultancy. Founded in 2005, we have 400+ practitioners spanning nine international offices, serving organisations across all regions and major sectors.

This guidance is general information based on a limited number of incidents and is not advice for any specific organisation. Indicators can be changed or copied. Take tailored advice before acting.