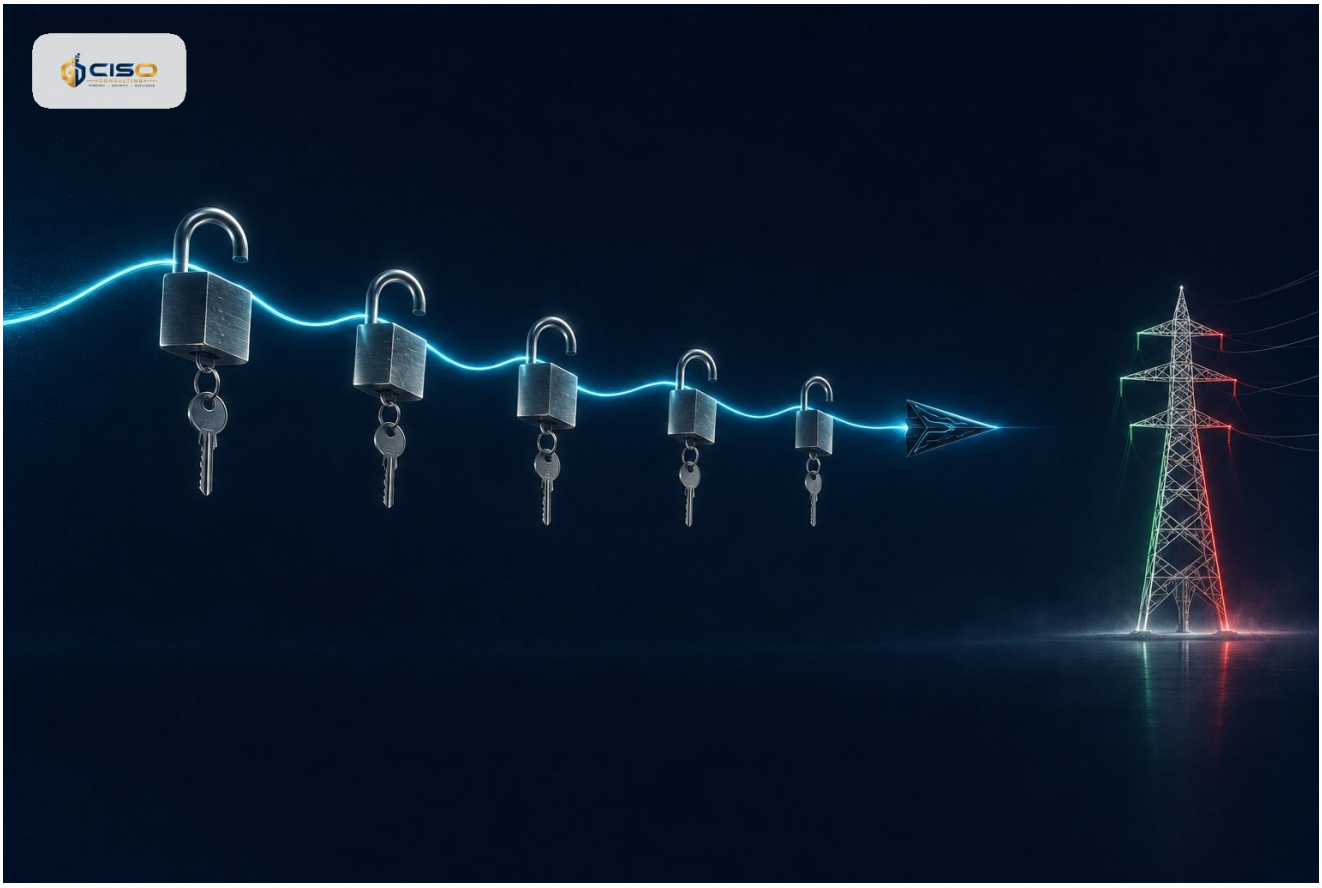
Italy and Switzerland warn of autonomous AI-agent attacks on critical infrastructure



Italian and Swiss agencies report AI agents compromising critical infrastructure organizations at speed by chaining known flaws and weak credentials. They urge fewer exposed systems, faster patching, MFA and monitoring.

- CSIRT Italia detected large-scale campaigns in which autonomous AI agents found and exploited vulnerabilities and misconfigurations on exposed endpoints.
- In one case, a coordinator directed five agents through stages from information gathering to exploitation and persistence, reaching internal networks.
- Switzerland's NCSC confirmed its first successful AI-assisted attacks, affecting organizations in at least two sectors.
- The agencies named four frameworks involved: ARTEX AI, CyberStrikeAI, DeepSeek Harness - Red

Team and PentAGI.

- NCSC says the number of publicly available AI agents has nearly doubled since October.

Cybersecurity agencies in Italy and Switzerland have warned that automated attacks driven by AI agents are now hitting organizations in critical infrastructure sectors.

What the agencies observed

CSIRT Italia reported large-scale campaigns in which autonomous agents identified and exploited vulnerabilities and misconfigurations on exposed endpoints. In some cases several agents worked in parallel and compromised many systems within hours. In one DeepSeek Harness setup, a coordinator assigned tasks to five agents covering reconnaissance through to exploitation and persistence. The agents chained flaws, misconfigurations and compromised credentials, using exposed systems as a foothold into internal networks. The agency also said they wrote malicious code in real time to evade endpoint defenses.

Switzerland's NCSC confirmed its first reports of successful AI-assisted attacks, affecting operators in at least two sectors. According to the agency, every known case ended in compromise of the victim's IT systems. Agents exploited weak passwords and known vulnerabilities on exposed endpoints and decided their next steps independently.

Recommended measures

The Italian agency advises prioritizing patches for internet-facing systems, especially flaws known to be exploited, and disabling unneeded remote management interfaces. It also recommends monitoring control plane operations, authentication sources, traffic to internal networks and outbound server traffic. The Swiss agency recommends patching critical vulnerabilities within 24 hours of release, adding MFA or passkeys, and enforcing least privilege.

The source does not report any Saudi incidents. The pattern, however, relies on exposure and basic hygiene gaps that are relevant to any regulated institution.

What to do now

1. Inventory internet-facing systems and remove or restrict any that are not required.
2. Disable remote management interfaces that are not needed, and put the rest behind MFA.
3. Patch critical and actively exploited vulnerabilities on exposed systems within 24 hours where feasible.
4. Enforce MFA or passkeys, strong passwords and least privilege on exposed and administrative accounts.
5. Monitor authentication sources, internal-bound traffic and outbound server traffic for unusual activity.
6. Include AI-agent-driven, fast multi-stage attacks in red-team exercises and incident response playbooks.