From Swiss Cyberattacks to Westminster: Who Is Accountable for AI Agents?

Switzerland’s first confirmed AI-agent-assisted intrusions and a UK parliamentary vote reveal distinct challenges for cybersecurity, oversight and cross-border cooperation.

Vladimir Tsakanyan, Ph.D. · Center for Cyber Diplomacy and International Security

November 24, 2026

Cover image: AI-generated conceptual illustration. It does not depict an actual incident, facility or participant.

Switzerland’s cyber authority reported on November 24 that it had confirmed its first cases of attackers using AI agents against organizations and critical infrastructure in the country. On the same day, British lawmakers voted on amendments to a cybersecurity bill that includes proposed powers concerning a different problem: the loss of oversight or control of an AI agent affecting a regulated system. The juxtaposition exposes a central test for cyber diplomacy: institutions must respond both to deliberate misuse and to failures of control, without treating them as the same event. [1] [3]

The Swiss warning describes actual compromises reported to the authorities; the British measures remain part of a bill and prospective regulations. Neither announcement establishes that the Swiss cases involved a British operator, a runaway system, or a state-sponsored campaign. Their policy connection is analytical: both show why faster machine-assisted activity makes accountability, evidence preservation and cross-border coordination more urgent. [1] [4] [5]

What Switzerland has—and has not—confirmed

The Swiss National Cyber Security Centre says that, in recent days, it received its first reports of attacks against organizations in Switzerland carried out with assistance from AI agents. It says organizations and critical infrastructure operators in at least two sectors were affected and that, in all attacks known to it in this set, the victims’ IT systems were successfully compromised. This is a statement about the cases reported to that agency, not a measured success rate for all AI-assisted attacks. [1]

The announcement does not identify the victims, specify the number of affected organizations, name the sectors, assign responsibility to a government, or describe a

particular interruption of essential services. It is also not a claim that Switzerland witnessed the world's first use of AI in a cyberattack. The distinction matters: an official warning can be significant without answering every question the public might reasonably have.

The Swiss authority describes agentic offensive frameworks as systems that can help search for weaknesses, select tools, evaluate results and plan subsequent actions. Such tools can also serve legitimate security testing. The presence of automation does not make an intrusion self-originating: a person or organization may still choose the target, configure the tool and use its results. Nor does an AI label by itself tell investigators how much of an operation was automated. [1]

Speed changes the defender's clock

A separate November 24 assessment from the UK's National Cyber Security Centre describes a rise in AI-assisted attacks it has observed since September. It reports rapid attempts against internet-facing assets, with some incidents causing service downtime or system compromise. Its central point is that familiar weaknesses, rather than a wholly new class of vulnerability, have often provided the opening. The UK and Swiss reports do not establish that they describe the same incidents or actors. [2]

Both agencies point to weaknesses such as exposed services, weak or default credentials and inadequate access controls. The Swiss authority also emphasizes known software vulnerabilities. The UK assessment discusses leaked credentials, configuration errors and weaknesses in application interfaces. The analytical consequence is not that ordinary security suddenly becomes irrelevant; it is that unaddressed weaknesses can be found and pursued with less time for human defenders to react. [1] [2]

That shift changes the economics of resilience. A forgotten remote-access service may once have survived because few people investigated it. A system that can repeatedly identify exposed assets, test common weaknesses and adapt to responses may reduce that accidental protection. The British agency notes that current tools can be noisy, which may help detection; it also cautions that future versions could place greater emphasis on stealth. A policy response should therefore measure detection and containment, rather than assume that speed always favors one side. [2]

Westminster's narrower question

In the House of Commons on November 24, members considered amendments sent from the House of Lords to the Cyber Security and Resilience (Network and Information Systems) Bill. Hansard records that the Commons agreed to Lords amendments 33 through 106 and 109 through 112. Amendment 33 addresses potential functions concerning a relevant network or information system compromised, or at risk of compromise, through loss of oversight or control of an AI agent. The Commons also rejected a proposed change to that

amendment by 278 votes to 93. [3] [8]

The legislation is not yet settled law. The Commons disagreed with two other Lords amendments, and the parliamentary process must continue; the record of a Commons vote is not Royal Assent. Moreover, the AI-agent provision is an enabling measure: it contemplates functions being conferred through regulations, including definitions to be set later. It should not be presented as an already operative, comprehensive AI safety regime or as a rule automatically covering every AI product. [3] [4] [9]

The government's November 3 factsheet describes possible future measures following consultation, including resilience requirements, notification of affected parties, preservation of evidence and expert post-incident reviews. These are proposals for secondary legislation, not requirements that operators can assume are already in force. The factsheet says the envisaged measures target loss-of-control incidents involving regulated entities and are not intended to establish a comprehensive framework for all AI harms. [5]

An attacker's agent is not a runaway agent

This distinction should shape every comparison between the two countries. In the Swiss cases, the authority says attackers used AI agents to assist intrusions. The UK provision concerns a compromise or risk that is or may be attributable to losing oversight or control of an AI agent. A malicious actor can direct a system effectively; an organization's own system can behave outside its intended boundaries without a malicious intruder. One incident might involve both, but that requires evidence, not a convenient narrative. [1] [4]

The safeguards overlap only partly. Defending against a hostile agent requires reducing exposed services, protecting credentials, limiting privileges and responding quickly to unusual activity. Governing an organization's own agent also requires clear authorization, boundaries on actions, meaningful monitoring, the ability to interrupt operations and records that allow a later investigation. A procurement contract should identify what the developer, service provider and operator can each observe and control. None of those responsibilities removes the attacker's responsibility for an intrusion.

Members of Parliament raised precisely this distribution of responsibility. During the debate, Dame Chi Onwurah questioned whether duties focused on those deploying agents would adequately address the role of the companies that design them. Other members argued for rules framed around potentially harmful autonomous systems more broadly, rather than around a particular technology. Those are competing policy arguments in a legislative debate, not findings that a named developer caused the Swiss intrusions. [3]

Reporting is infrastructure, too

Switzerland has required covered critical-infrastructure operators to report qualifying cyberattacks to its national authority within 24 hours of discovery since April 1, 2025. That is

an existing national obligation with a defined scope, not a new universal reporting duty created by the November 24 warning. Rapid initial notification can alert responders before an investigation is complete, but the first report may contain uncertainty about cause, impact and attribution. [6]

The European Union's NIS2 framework offers a related, distinct model for significant incidents affecting covered entities: an early warning within 24 hours of awareness, a notification within 72 hours and a final report within the prescribed period. Switzerland is not an EU member, and the UK bill is not NIS2. The regimes have different legal bases, coverage, thresholds and authorities. The useful comparison is institutional: each must turn incomplete early signals into reliable decisions without encouraging false certainty. [7]

For cross-border operators, the first practical question is who receives what information and when. A cloud provider, a hospital, a regulator and a national response team may each see a different part of the same disruption. They also have different legal obligations and reasons to protect sensitive material. Common incident categories, secure channels, points of contact and agreed correction procedures can reduce duplication without pretending that every jurisdiction has identical rules.

The diplomatic value of a bounded disclosure

The UK bill's Lords amendments include restrictions on certain disclosures of information by network-and-information-system enforcement authorities to foreign law-enforcement authorities. The specified jurisdictions include Switzerland, alongside Australia, Canada, European Economic Area states, New Zealand and the United States. This is a narrowly framed provision in a bill, not proof of a new UK–Swiss agreement and not a blanket rule governing every form of technical cooperation. [8]

Its significance is the balance it illustrates. Incident information can help partners recognize shared infrastructure, tactics and vulnerabilities; indiscriminate disclosure can expose commercially sensitive data, personal information or investigative methods. Effective cyber diplomacy needs both timely exchange and credible limits on purpose, access and subsequent use. Governments should also avoid assuming that a technical indicator alone establishes the identity or legal responsibility of an attacker.

The Swiss authority says it is in contact with national and international partners and shares technical indicators with critical-infrastructure operators. That is an important operational statement, but it does not identify a common campaign with the United Kingdom or establish which information was shared with which foreign government. The diplomatic opportunity lies in compatible processes for testing, protecting and correcting information across borders—not in manufacturing a connection that the evidence does not show. [1]

From disclosure to measurable resilience

For policymakers, three tests follow. First, can covered organizations identify exposed systems, protect credentials and apply appropriate security updates at the speed the threat now demands? Swiss guidance recommends patching critical updates on internet-facing systems as soon as possible and no later than 24 hours after release; that is the agency's recommendation, not a statement that every organization can safely patch every system instantly. Operators of safety-critical equipment still need tested procedures, compensating controls and recovery plans. [1]

Second, can authorities reconstruct what happened? Preserving relevant logs, time stamps, permissions and changes to system configuration can help distinguish a human-directed intrusion from malfunction, misuse of a legitimate account or loss of control of an authorized agent. Records should be proportionate and protected: logging more sensitive information than necessary can create another security problem. Proposed British evidence-preservation measures will need to be judged by their eventual scope and safeguards. [5]

Third, can institutions learn across borders without turning uncertainty into attribution? An initial alert should communicate what has been observed, what remains unverified and how recipients can act. A later assessment can refine the facts. Exercises that test notification, evidence handling and recovery across suppliers and jurisdictions would make those commitments more concrete than another declaration alone.

The November 24 developments are a warning about institutional speed as much as machine speed. Switzerland has identified a real set of reported compromises; Britain is debating how its law should respond to a separate class of AI-related risks. The strongest response is to keep those categories clear, invest in the unglamorous controls that make essential services resilient, and build trusted channels for sharing evidence before an incident becomes a diplomatic crisis. [1] [3]

Sources and links

[1] Swiss National Cyber Security Centre: Cyberattacks involving AI agents—first confirmed cases in Switzerland (24 November 2026)

https://www.bacs.admin.ch/en/26-ki-agenten-1-en

[2] UK National Cyber Security Centre: Familiar weaknesses, exploited faster (24 November 2026)

https://www.ncsc.gov.uk/blogs/familiar-weaknesses-exploited-faster-defending-against-agentic-cyber-attacks

[3] UK Parliament, Hansard: Cyber Security and Resilience Bill, Commons consideration of Lords amendments (24 November 2026)

https://hansard.parliament.uk/Commons/2026-11-24/debates/9696CEE9-A8C3-449D-9B15-89FA9F00CF8E/CyberSecurityAndResilience(NetworkAndInformationSystems)Bill

[4] UK Parliament: Cyber Security and Resilience Bill, text including proposed AI-agent provisions

https://bills.parliament.uk/publications/68323/documents/8980

[5] UK Department for Science, Innovation and Technology: Agentic AI resilience measures (updated 3 November 2026)

https://www.gov.uk/government/publications/cyber-security-and-resilience-network-and-information-systems-bill-factsheets/agentic-ai-resilience-measures

[6] Swiss National Cyber Security Centre: Critical infrastructure cyberattack reporting obligation (31 March 2025)

https://www.bacs.admin.ch/en/25-mp2-en

[7] European Commission: NIS2 Directive—frequently asked questions and incident reporting

https://digital-strategy.ec.europa.eu/en/faqs/directive-measures-high-common-level-cybersecurity-across-union-nis2-directive-faqs

[8] UK Parliament: Cyber Security and Resilience Bill, Lords amendments (Bill 175, 3 November 2026)

https://publications.parliament.uk/pa/bills/cbill/59-02/0175/260175.pdf

[9] UK Parliament: Cyber Security and Resilience Bill—current passage

https://bills.parliament.uk/bills/4035

Read the original article online

https://cybercenter.space/insights/switzerland-ai-agent-attacks-uk-cybersecurity-bill-2026/