cyberaci

AI IN CRITICAL INFRASTRUCTURE: SECURITY AND RESILIENCE

When AI becomes the attack surface

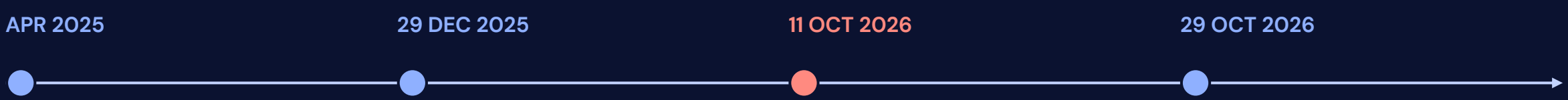
18 months of attacks. AI on both sides. What NIS2 and standards can fix.

Asif Safdary Founder & CEO, CyberACI

asif.safdary@cyberaci.com | cyberaci.com

FOUR INCIDENTS, 18 MONTHS. NO AI. NO ZERO-DAY.

WHAT ACTUALLY HAPPENED



Norway, dam[1]

Valve open 4 h, unnoticed

Web panel, weak password

Poland, grid edge[2]

30+ substations, 1 CHP plant

Exposed VPN, no MFA, wiper

France, 700 stores[3]

Controllers on public IPs

~20 fires, hydro scanned too

Binance, $6.36 bn[4]

Gone in ~20 minutes

Signing infrastructure compromised

Entry: remote access on a public IP. Every time.

1 Norwegian Police Security Service (PST), Aug 2025 2 CERT Polska incident report, Jan 2026
3 CERT-FR CERTFR-2026-ALE-012, 12 Oct 2026 4 TRM Labs, 29 Oct 2026

cyberaci

BINANCE: 36 HOURS OF STAGING, ~20 MINUTES OF THEFT

ANATOMY OF THE LARGEST HEIST ON RECORD



$6.36 bn
10 chains[1]

~20 min
first to largest sweep[1]

36 h
staging, visible on-chain[1]

37 min
to the first freeze[2]

The detection window was the 36 hours.

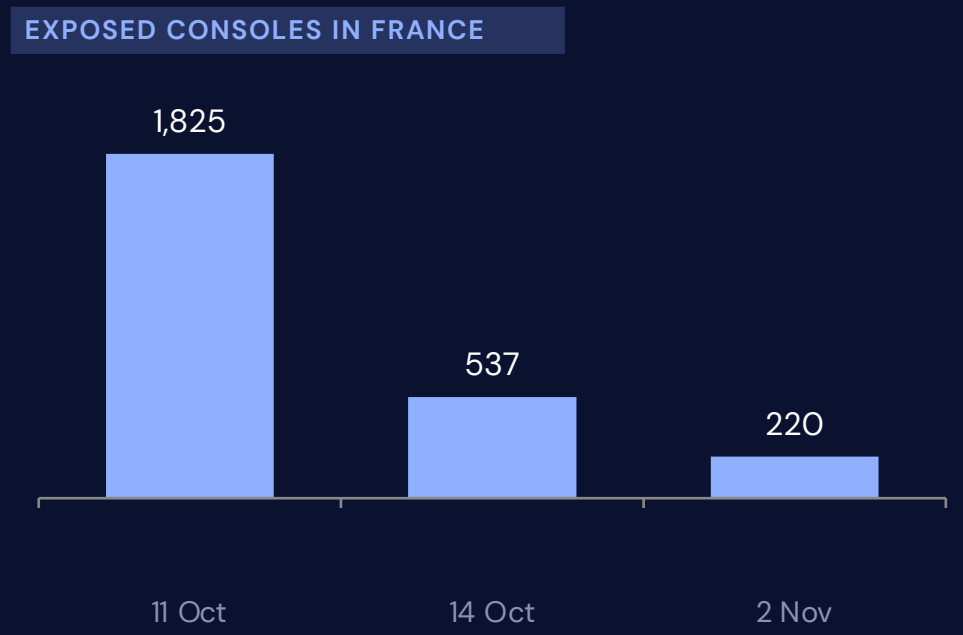
1 TRM Labs, 29 Oct 2026
2 Hacken, Binance hack analysis, Oct 2026

cyberaci

FRANCE: 700 SITES SABOTAGED THROUGH FRIDGE CONTROLLERS

OT THAT NOBODY CALLS OT

DEVICE	Store controller with a web console, run by the fridge contractor
ACCESS	Public IP, end-of-life firmware
ACTION	Defrost on, alarms off, setpoints to 50 °C[1]
IMPACT	~20 fires[2], tens of millions in a day
ALSO	Hydro plants scanned, Modbus 502[2]
WHO	IRGC unit Shahid Kaveh, attributed 9 Nov[2]



Three weeks to unplug. Nothing fixed inside.

The contractor owned the access. The retailer owned the fires.

1 CERT-FR alert CERTFR-2026-ALE-012, 12–16 Oct 2026
2 French government attribution to the IRGC, reported by The Hacker News, Nov 2026

NOT A TECHNOLOGY PROBLEM. A TIME AND CONTROL PROBLEM.

THE PATTERN

Got in	Noticed	Lost
Exposed remote access Weak or no MFA	Hours later Or by the fire	Control of a physical process

~20 min

Binance: to move $6.36 bn[1]

4 h

Norway: valve open, unnoticed[2]

24 h

NIS2: early warning deadline[3]

1 TRM Labs, 29 Oct 2026 2 Norwegian Police Security Service (PST), Aug 2025
3 Directive (EU) 2022/2555 (NIS2), Art. 23(4)

cyberaci

SIX WAYS THE AI ITSELF BECOMES THE ATTACK SURFACE

NEW CYBER RISKS FROM AI IN CRITICAL INFRASTRUCTURE

1 **Prompt injection**

via your own logs

2 **Poisoned feedback**

it learns what you close

3 **Model supply chain**

whose weights?

4 **Dependency**

June 2026: 3 weeks without frontier models[1]

5 **Economic DoS**

pay per alert

6 **Autonomy**

an agent with OT write access

1 US Department of Commerce order of 12 Jun 2026 suspended two Anthropic frontier models; access restored 1 Jul 2026 (IAPP, 15 Jun 2026; Anthropic)

cyberaci

NISG 2026: IN FORCE SINCE 1 OCTOBER. SILENT ON AI.

THE REGULATORY FRAME

AUSTRIA[1]

1 Oct 2026	duties apply
31 Dec 2026	register with the BCS
1 Oct 2027	self-declaration
1 Oct 2028	audits on request
§ 31	management is liable

REPORTING[2]

24 h
early warning

72 h
notification

1 mo
final report

ART. 21(2): TEN MEASURES[2]

a Risk analysis	**b Incident handling**
c Continuity	**d Supply chain**
e Vulnerabilities	**f Effectiveness**
g Hygiene, training	h Cryptography
i Access, assets	**j MFA**

Blue: failed in every incident on slide 2

~4,000 Austrian organisations[3]

1 NISG 2026, BGBl. I Nr. 94/2025
2 Directive (EU) 2022/2555 (NIS2), Art. 21 and 23
3 Austrian Economic Chamber (WKO) estimate

AI DOES NOT MAKE YOU COMPLIANT. IT MAKES FOUR MEASURES PROVABLE.

WHERE AI ACTUALLY HELPS WITH ART. 21

b	Incident handling	**core**	evidence in hours, not days
f	Effectiveness	**core**	measured time-to-detect
i	Access, assets	**core**	finds the exposed controller
a	Risk analysis	**supports**	live exposure picture
e	Vulnerabilities	**supports**	prioritise by real exposure
d	Supply chain	**supports**	and the AI vendor is one
j	MFA	**detects**	flags gaps, enforces nothing
c	Continuity	**detects**	no model restores a relay
g, h	Training, crypto	**none**	people and architecture

MINERVA: AI ANALYSTS INSIDE THE OPERATOR'S BOUNDARY



26,211
alerts in 7 days, 1,272 investigated

1,156
investigations closed by agents

116
escalated to a human (9%)

1,156 closed by agents × 30 min of analyst time ≈ 578 hours ≈ **72 person-days saved** in one week

cyberaci

WHAT HELD UP, AND WHAT WE HAVE NOT SOLVED

LESSONS LEARNED, INCLUDING THE UNSOLVED ONES

HELD UP

Evidence beats opinion

analysts trust what they can re-run

Boundaries in code

not in a system prompt

Review is training

confirm or overturn becomes a fact

UNSOLVED

Certifying what learns weekly

AI Act standards still missing

Who audits the operator's rules?

a bad directive has authority

The sovereignty tax

own GPU trails the cloud frontier

cyberaci

STANDARDISE CONTROL POINTS AND EVIDENCE. NOT THE TECHNOLOGY.

WHAT STANDARDS SHOULD FIX, FROM A BUILDER'S CHAIR

1 **Control points**

where a human must approve

2 **Evidence**

every AI verdict auditable

3 **Sovereign deployability**

runs inside the boundary, keeps running

Resilience = how fast you see × how much you still control

AI helps when it shortens the first and leaves the second with the operator.

cyberaci

cyberaci

asif.safdary@cyberaci.com | cyberaci.com

When AI becomes the attack surface



Asif Safdary Founder & CEO, CyberACI