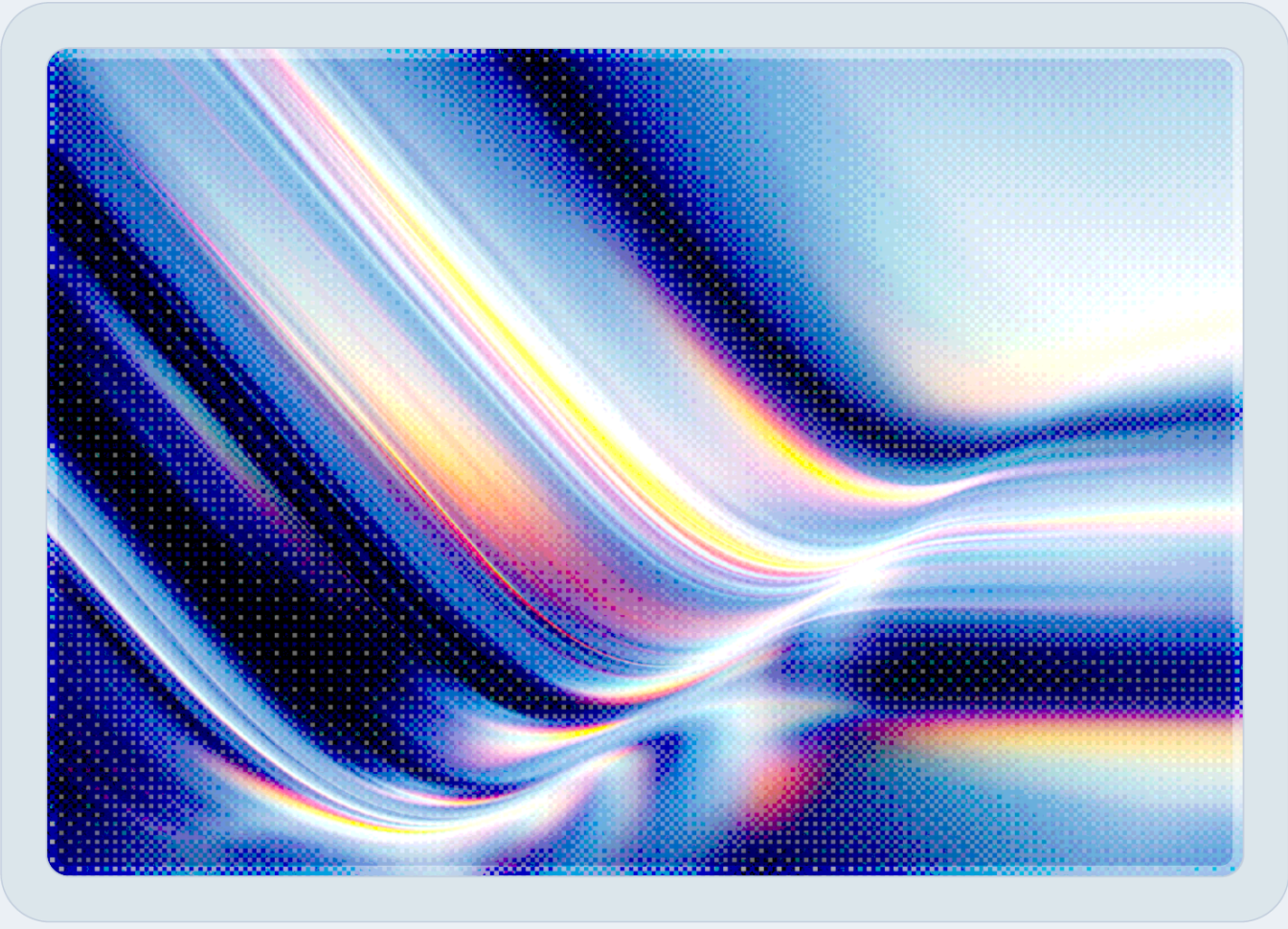
CSAI Foundation | Cloud Security Alliance

Proving Oversight: NIS2 Board Evidence for AI Agent Access

Delegation chains, third-party agents, and incident reporting under Article 20

2026-11-05

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

Article 20(1) of the NIS2 Directive requires the management bodies of essential and important entities to approve the cybersecurity risk-management measures taken under Article 21, to oversee their implementation, and to be capable of being held liable for infringements of that Article [1]. CSA's companion note of 4 November 2026 argued that decisions about what AI agents may access are access control decisions under Article 21(2)(i), and therefore fall within what the board approves [8]. This note takes that conclusion as its starting point and asks a narrower, more practical question. If a supervisor or a court later asks whether a board *oversaw* the access of its non-human workforce, what evidence would answer that question, and where is it most likely to be missing?

Our analysis identifies three areas where agent deployments create evidence gaps that conventional access reviews are unlikely to close. The first is delegation. When an agent acts on behalf of a user, or calls another agent, the chain of accountability that Commission Implementing Regulation (EU) 2024/2690 assumes for identities tends to break down [2][7]. The second is third-party agents and agent tooling, which bring vendor-operated identities into the entity's systems under Article 21(2)(d) supply chain obligations [1]. The third is incident reporting. An entity that cannot separate agent activity from human activity may struggle to meet the 24-hour, 72-hour, and one-month reporting cycle in Article 23 [1]. CSA survey data indicates that 68% of organizations cannot clearly distinguish AI agent activity from human activity, and that ownership of agent identity is spread across security, engineering, IT, and IAM teams [6].

The practical recommendation is that management bodies ask for a standing "agent access evidence pack". This is a short, recurring board report that records approval, review, and exceptions for agent identities, delegation patterns, third-party agent access, and incident readiness. Without such a record, an entity may have reasonable controls but little documentary proof that the board oversaw them. This reading is an interpretation of the Directive's text. No supervisory authority has, to our knowledge, issued guidance or an enforcement decision linking Article 20 specifically to AI agents.

Background

NIS2 (Directive (EU) 2022/2555) required Member States to transpose its provisions by 17 October 2024 [1]. Article 20 sets out the governance duty. Paragraph 1 places approval and oversight of Article 21 measures with the management body and requires that the management body can be held liable for

infringements. Paragraph 2 requires members of management bodies to follow training so that they can identify risks and assess cybersecurity risk-management practices [1]. Article 21 then defines the measures themselves. These include supply chain security (point d), "human resources security, access control policies and asset management" (point i), and multi-factor or continuous authentication where appropriate (point j) [1]. Article 21(3) adds that, when considering supply chain measures, entities must take into account the vulnerabilities specific to each direct supplier and the quality of suppliers' cybersecurity practices [1].

The enforcement architecture reaches individuals as well as entities. For essential entities, Article 32(5) permits authorities, where earlier measures have been ineffective, to suspend a certification or authorisation concerning the entity's services or to seek a temporary prohibition on a person exercising managerial functions at chief executive or legal representative level. Article 32(6) requires that natural persons responsible for or acting as legal representatives of an essential entity can be held liable for breaching their duty to ensure compliance [1]. Article 34 sets the floor for national maximum fines at no less than EUR 10 million or 2% of worldwide annual turnover, whichever is higher, for essential entities, and EUR 7 million or 1.4%, whichever is higher, for important entities [1]. These provisions take effect through national law, which continues to vary. Germany's revised BSI Act, for example, obliges management in § 38 to implement and oversee risk-management measures, makes management liable to the entity under the company-law rules for its legal form, and requires regular training [4]. Transposition elsewhere remained incomplete as of mid-2026. On 8 July 2026 the Commission referred Ireland, Spain, France, and the Netherlands to the Court of Justice of the EU for failing to notify full transposition [5].

For a defined list of digital entities, including cloud computing, data centre, managed service, and managed security service providers, DNS and TLD operators, online platforms, and trust service providers, Implementing Regulation (EU) 2024/2690 makes Article 21 concrete [2]. Section 11 of its annex covers access control. Section 11.5 requires entities to manage the full life cycle of identities of network and information systems and their users, to link user identities to a single person, and to permit shared identities only where necessary for business or operational reasons and subject to explicit approval and documentation [2]. ENISA's NIS2 Technical Implementation Guidance, published in June 2025, pairs these requirements with examples of documentary evidence that entities can produce [3]. That emphasis on evidence is central to this note. We think Article 20 oversight is likely to be assessed in a similar way: a supervisor asking whether the board oversaw a measure will look for a documented, reviewable record of approval and review. The Directive does not say this explicitly, and the analogy is CSA's interpretation rather than a requirement stated in the guidance.

Security Analysis

The three gaps introduced above share a common root. Each arises where agent activity cannot be traced cleanly to an approved identity, an accountable person, or a documented decision, and each therefore weakens the record a board would rely on to demonstrate oversight. The subsections that follow examine delegation, third-party agents, and incident attribution in turn, then bring them together in a single mapping of obligations to evidence.

Delegation chains break the accountability link

The Implementing Regulation's identification model assumes that each identity is unique and that user identities trace back to one person [2]. Agents complicate that assumption in a specific way. An agent often acts *on behalf of* a human user, inheriting or exchanging that user's authority, and it may in turn invoke tools, other agents, or third-party APIs. CSA's research note on non-human identity management for agentic AI observes that where this delegation chain is not preserved cryptographically, "every step downstream of the original prompt loses provenance" [7]. Standards to preserve it do exist. OAuth 2.0 Token Exchange (RFC 8693) defines an "act" claim that keeps the acting party distinct from the party on whose behalf it acts [11]. CSA's analysis, however, characterizes adoption of such mechanisms in agent runtimes as uneven [7].

Survey data shows that patterns which do not preserve this distinction are widespread, even if they are not universal. In CSA's January 2026 survey of 228 IT and security professionals, 52% of organizations reported using workload identities for agents, but 43% also reported using shared or generic service accounts and 31% reported running agents under a human user's identity. Respondents could select more than one option [6]. In the shared-account case, the agent's actions trace to no single accountable person. In the borrowed-identity case, they trace to the wrong one. Either way, the board-approved access control policy may be satisfied on paper while the identity record in the logs does not reflect who, or what, acted. For Article 20 purposes, this matters because oversight evidence normally relies on that identity record. A quarterly access review that lists a service account and its entitlements says little about which agents use it, under whose instruction, and for what purpose.

The governance consequence is that delegation should be treated as a policy matter for the board to approve, not an engineering detail. The board approves the policy that defines permitted delegation patterns, while maintaining the register of those patterns can sit with management. The approved policy should state which delegation patterns are permitted, such as agents acting with a user's delegated and scoped authority, agents running under dedicated workload identities, and agent-to-agent calls, and what provenance must be retained for each. Exceptions, such as a legacy integration that can only run under a shared account, map naturally onto the explicit approval and documentation route the Implementing

Regulation provides for shared identities [2]. A documented exception approved through a defined process is evidence of oversight. An undocumented shared account, by contrast, leaves no record that oversight occurred.

Third-party agents widen the supply chain perimeter

Much agent activity inside an entity's systems is performed by agents the entity did not build. SaaS vendors embed agents in their products, managed service providers run agents against customer environments, and developers connect internal systems to externally maintained tool servers. Each of these arrangements introduces identities, tokens, or delegated grants that the entity's access control policy must cover, but that are configured and operated in part by someone else. Article 21(2)(d) and Article 21(3) bring this into the scope of supply chain security, requiring entities to consider supplier-specific vulnerabilities and the quality of suppliers' cybersecurity practices [1].

Practitioners see the credential risk in agent tooling clearly. In CSA's survey, 79% of respondents agreed that agents introduce new access pathways that are difficult to monitor, and 74% agreed that agents often receive more access than necessary [6]. These figures measure perception rather than incidents or exposure. OWASP's Non-Human Identities Top 10 for 2025 independently ranks improper offboarding and secret leakage among the leading risks for machine identities [12]. Both become harder to manage when the identity sits on a vendor's platform. Traditional supplier questionnaires, which tend to focus on the vendor's own infrastructure, may not ask how vendor agents hold and use credentials inside the customer's environment. A board that approves a supplier security policy without asking how vendor-operated agents authenticate, what scope they hold, and how quickly the entity can revoke them may therefore be approving a policy that does not reach the access that matters most.

The evidence a board should expect here is specific. It includes an inventory of third-party agents and tool integrations with production access, the grant or credential type each uses, contractual terms on how access is suspended or revoked, and the date of the last review. CSA's research on shadow AI agents frames the discovery and inventory of agents, including those deployed without governance oversight, as an identity problem best addressed at the non-human identity perimeter [14]. CSA's AI Act–NIS2 conformity analysis further notes that supplier attestations, such as AI-CAIQ responses and STAR for AI registrations, can serve as shared evidence across regimes [9]. That approach applies equally to agent providers.

Incident reporting depends on attribution

Article 23 requires essential and important entities to submit an early warning within 24 hours of becoming aware of a significant incident, an incident notification within 72 hours, and a final report within one month, with intermediate reports on request and a progress report if the incident is still ongoing at that point [1]. For entities covered by the Implementing Regulation, the criteria for a significant incident include a successful, suspectedly malicious and unauthorised access to network and information systems that is capable of causing severe operational disruption [2][13]. An agent that is manipulated into misusing legitimate access may meet that description, depending on the circumstances. The difficulty is establishing quickly that this is what happened.

Here the attribution gap becomes a reporting risk. If 68% of organizations cannot clearly distinguish agent activity from human activity [6], many will begin an investigation unsure whether an employee, a compromised account, or an agent acting on manipulated input caused the event. When an agent runs under an employee's credentials, the initial evidence points at the employee. That ambiguity can delay the 24-hour early warning and distort the 72-hour assessment of severity and indicators of compromise. A later finding that the entity lacked the logging needed to attribute agent actions could also bear on whether the board oversaw the measures Article 21 requires, since incident handling and logging are themselves listed Article 21 measures [1].

For the board, the relevant evidence is whether agent identities are distinguishable in logs, whether the incident response plan contains an agent-specific playbook, and whether revocation of an agent's credentials has been tested. CSA's survey found that disabling identities or revoking tokens is a common containment action for agents [6]. That action is only reliable when each agent has its own revocable identity, which is also the precondition for the discovery and attribution work CSA's shadow-agent research describes [14].

Summary: obligations, agent-specific gaps, and board evidence

The table below brings the three areas together with the core access control obligation. It pairs each NIS2 provision with the agent-specific gap this analysis identifies and the evidence a management body could reasonably ask to see. The evidence column reflects CSA's interpretation and is not regulatory text.

NIS2 provision	Agent-specific gap	Board-level evidence of oversight
Art. 21(2)(i) access control; CIR §11.2 rights management [1][2]	Agents inherit broad human or shared entitlements; 74% of survey respondents agree agents often receive more access than necessary [6]	Approved policy that names agents explicitly; periodic review results for agent entitlements
CIR §11.5 identification [2]	Delegation chains and shared accounts break the link to an accountable person [6][7]	Register of permitted delegation patterns; approved and documented exceptions for shared identities
Art. 21(2)(d), 21(3) supply chain [1]	Vendor-operated agents and tool integrations hold credentials the entity does not fully control	Third-party agent inventory; contractual revocation terms; supplier attestations [9]
Art. 23 incident reporting [1]	Inability to attribute agent actions delays the 24h/72h cycle	Agent tagging in logs; agent incident playbook; tested revocation time
Art. 20(2) training [1]	Training curricula may omit delegated machine autonomy	Training records showing agent risk covered in the management curriculum
Art. 20(1) oversight [1]	Fragmented ownership; agent identity often has no single accountable function [6]	Named accountable owner; recurring agent access report in board minutes

Taken together, the rows describe the contents of the agent access evidence pack recommended in this note, and they form the basis for the recommendations that follow.

Limits of this analysis

The analysis should be read with three caveats. First, the Implementing Regulation binds only the entity types it lists. For other essential and important entities, it may inform what supervisors regard as "appropriate and proportionate", but it is not directly applicable [2]. Second, CSA's survey covers

organizations generally, not NIS2-regulated entities specifically, so its figures indicate likely patterns rather than measured compliance rates [6]. Third, the scope of personal liability depends on national law. Germany, for instance, routes management liability to the entity through company law rather than creating direct liability to third parties [4]. We expect the evidence-based approach recommended here to remain useful across these variations, because a documented record of oversight is relevant whether the question comes from a supervisor, an auditor, an insurer, or a shareholder.

Recommendations

The recommendations below are sequenced by time horizon. They move from establishing a baseline record, through extending existing controls to agents, to the structural changes that make oversight sustainable. Throughout, the depth of effort should match the entity's actual agent footprint, consistent with Article 21's requirement that measures be appropriate and proportionate.

Immediate Actions

Management bodies in scope should place agent access on the board agenda as a defined item and commission the first version of an agent access evidence pack. At minimum, it should address the six rows of the table above. The depth of the pack should match the entity's actual agent footprint, and an entity with limited, low-privilege agent use may be able to cover several rows in a single paragraph. The CISO, or whoever the board designates, should confirm in writing whether the currently approved access control policy covers AI agents, delegated authority, and third-party agents. If it does not, the CISO should bring an amendment for approval so that the minutes record the decision.

In parallel, security teams should identify every production agent running under a named human's credentials or a shared service account. For each, they should either migrate it to a dedicated identity or record it as an approved exception with a business justification, consistent with the Implementing Regulation's treatment of shared identities [2]. Teams should also confirm that they can revoke any production agent's access within a defined period. The measured time should be reported to the board, so that the figure reflects a test rather than an assumption.

Short-Term Mitigations

Over the next one to two board reporting cycles, organizations should define permitted delegation patterns and require that agent actions retain provenance, meaning the agent identity, the delegating principal, and the task context. Token exchange mechanisms such as the RFC 8693 "act" claim are one

route where platforms support them [11]. Where they do not, compensating logging should record the delegating principal alongside the agent identity. Agent entitlements should join the same review cycle as privileged human access.

Supplier management should extend to agents in the same period. Procurement and third-party risk teams should add agent-specific questions to supplier assessments covering how vendor agents authenticate, what scope they hold, where credentials are stored, and how access is suspended. They should also seek contractual commitments on revocation and incident notification. Alongside this, the incident response plan should gain an agent-specific playbook that addresses attribution, credential revocation, and the Article 23 reporting decision. The next round of Article 20(2) training should then cover delegated autonomy, prompt-driven misuse of legitimate access, and the board's role in approving delegation policy.

Strategic Considerations

Over the longer term, entities should assign a single accountable owner for non-human and agent identity. Article 20 concentrates accountability at the top, and fragmented ownership below it makes that accountability harder to evidence [6]. Organizations operating in several Member States should track national transposition and supervisory guidance, because the mechanics of individual liability differ [4] [5]. Entities that are also subject to the EU AI Act should design the evidence pack so it can be reused across both regimes, an approach CSA's conformity-gap analysis recommends more generally [9]. Finally, boards should expect this area to change. Agent identity standards, platform support for delegation provenance, and supervisory expectations are all developing, and revisiting the evidence pack on a fixed cadence is itself a form of oversight that can be documented.

CSA Resource Alignment

This note is a companion to CSA's research note **NIS2 Article 20: Board Liability for AI Agent Access Decisions** [8], published on 4 November 2026. That note establishes why agent access falls within the Article 21 measures that management bodies approve, and it compares Implementing Regulation requirements with observed agent practice. The present note builds on that foundation. It concentrates on the evidence that demonstrates oversight, and on three areas the earlier note treated briefly: delegation chains, third-party agents, and incident attribution. Readers new to the topic should start with the earlier note.

CSA's survey report **Identity and Access Gaps in the Age of Autonomous AI** [6] supplies the empirical basis for the gaps identified here, including shared and borrowed identities, weak attribution, fragmented ownership, and perceived over-provisioning. Its finding that organizations rely on governance stopgaps in place of identity-centric controls is directly relevant to Article 20. Procedural safeguards count as oversight evidence only if the board actually reviews them. Building on that data, CSA's research note **Detecting Shadow AI Agents at the Non-Human Identity Perimeter** [14] addresses how entities discover agents they did not know were running, which is the starting point for the agent and third-party inventories the evidence pack requires.

CSA's research note **Non-Human Identity Management for Agentic AI: Extending IAM Beyond Humans** [7] provides the technical grounding for the delegation analysis. It frames agent identity as an extension of existing IAM discipline and identifies loss of provenance across delegation chains as a key difference from earlier non-human identity categories. Its discussion of token exchange and workload identity gives engineering teams concrete primitives for producing the provenance evidence that boards should request.

AI Act Meets NIS2: The Conformity Gap in Critical Sector AI [9] addresses the multi-regime context in which many NIS2 entities operate. It recommends formal accountability delegation and shared control evidence, both of which the agent access evidence pack is designed to support. Underpinning all of these, the **AI Controls Matrix (AICM) v1.1** [10] offers a control framework whose Identity and Access Management; Supply Chain Management, Transparency, and Accountability; and Governance, Risk and Compliance domains can be mapped to Articles 20, 21, and 23. Its shared-responsibility model across model providers, orchestrated service providers, application providers, and AI customers helps allocate responsibility for third-party agent access.

References

[1] European Parliament and Council of the European Union. "Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2 Directive)." Official Journal of the European Union, December 2022.

[2] European Commission. "Commission Implementing Regulation (EU) 2024/2690 laying down rules for the application of Directive (EU) 2022/2555 as regards technical and methodological requirements of cybersecurity risk-management measures." Official Journal of the European Union, October 2024.

[3] ENISA. "NIS2 Technical Implementation Guidance." European Union Agency for Cybersecurity, June 2025.

[4] Federal Republic of Germany. "§ 38 BSIG – Umsetzungs-, Überwachungs- und Schulungspflicht für Geschäftsleitungen." Gesetze im Internet, Bundesministerium der Justiz, accessed November 2026.

[5] European Commission. "Commission refers Ireland, Spain, France and the Netherlands to the Court of Justice for failing to transpose rules on cybersecurity." Shaping Europe's Digital Future, July 2026.

[6] Cloud Security Alliance. "Identity and Access Gaps in the Age of Autonomous AI." Cloud Security Alliance, March 2026.

[7] Cloud Security Alliance. "Non-Human Identity Management for Agentic AI: Extending IAM Beyond Humans." CSA AI Safety Initiative, June 2026.

[8] Cloud Security Alliance. "NIS2 Article 20: Board Liability for AI Agent Access Decisions." CSA AI Safety Initiative, November 2026.

[9] Cloud Security Alliance. "AI Act Meets NIS2: The Conformity Gap in Critical Sector AI." CSA AI Safety Initiative, June 2026.

[10] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." Cloud Security Alliance, 2025.

[11] Jones, M., Nadalin, A., Campbell, B., Bradley, J., and Mortimore, C. "RFC 8693: OAuth 2.0 Token Exchange." Internet Engineering Task Force, January 2020.

[12] OWASP Foundation. "OWASP Non-Human Identities Top 10 – 2025." OWASP, 2025.

[13] Advisera. "Commission Implementing Regulation (EU) 2024/2690: Significant Incidents." Advisera, accessed November 2026.

[14] Cloud Security Alliance. "Detecting Shadow AI Agents at the Non-Human Identity Perimeter." CSAI Foundation, May 2026.