

Accountability for AI agents: from code signing to evidence

Amin Hasbini and Alan Seow · October 2026 · regulatory status as of 8 October 2026

Abstract

Organisations are deploying AI agents faster than they can account for what those agents do. Making a person answer for a group of agents often produces accountability theatre: responsibility without control. Drawing on two scenarios from practice, we argue that accountability is first a property of evidence, and set out what should be signed and recorded about agents, and how to keep those records verifiable through the move to post-quantum signatures, under rules that already apply in Europe. We disagree on how many people should supervise how many agents, and explain why. Three regional readings close the paper: Southeast Asia, the Gulf and Jordan, and the European Union.

1. Introduction

Artificial intelligence (AI)-powered organisations require two essential components: financial investment to support AI model and data centre subscriptions, and employees with strong coding expertise combined with a solid understanding of information technology (IT) environments. However, in Alan's cybersecurity experience, some organisations in Southeast Asia demonstrate a willingness to adopt AI-driven operations while neglecting critical investments in AI security governance. This imbalance creates significant risks, as technological advancement without robust governance undermines both resilience and accountability.

AI agents now act with real permissions: they raise purchase orders, change configurations and hand work to other agents. Adoption is outrunning governance, and accountability is being assigned to people rather than built into systems. This paper looks at that gap from practice. Two scenarios, one in an AI-driven organisation and one in an operational technology plant, show where assigned accountability breaks. An evidence model then shows what organisations can build instead, anchored in rules that already apply in Europe: the Digital Operational Resilience Act (Regulation (EU) 2022/2554, DORA) [1] for financial entities and the NIS2 directive (Directive (EU) 2022/2555) [2] for essential and important entities, where it has been transposed. We keep one disagreement open, on how many people should supervise how many agents, because the answer depends on the regulation around the organisation.

2. The “AI Superman”: accountability theatre

Scenario 1: as an illustration, in an AI-driven organisation, management may adopt a model where one human supervises ten AI agents, an approach we call the “AI Superman”. In this structure, the designated human bears responsibility whenever the AI agents malfunction. However, this arrangement effectively positions the “AI Superman” as a convenient shield for the Chief Information Security Officer (CISO) or senior management, deflecting accountability away from leadership.

In practice, consider an organisation with a software workforce: one human accounts for multiple AI agents acting as specialised workers, with some agents coordinating others, which we call child agents. The justification is to automate routine activities and to move the human up to monitoring acceptable boundaries and intervening when something goes wrong. Hence, it is not simply a matter of inserting a human: it requires sufficient visibility, authority, response time and organisational support.

If the designated human cannot meaningfully observe or intervene in what the AI agents did, then “human responsibility” is nominal. The human has responsibility without corresponding control.

One example is a procurement process powered by AI agents. Procurement is multi-task work, so several agents are deployed: a procurement agent, a bid evaluation agent and an invoicing agent, each with its associated child agents. It is not a one-off configuration. An AI agent can hallucinate and go sideways, so human intervention at both the verification and the validation stages is crucial.

3. Ground truth: AI agents in an OT plant

Scenario 2: in an operational technology (OT) environment, such as a power plant with machine valves, turbines and steam systems powering generators, strict safeguards are in place. These include network segmentation between corporate IT and production systems, least-privilege access, physical access controls, removable media restrictions and one-way data transfer mechanisms. Deploying large language models (LLMs) in such settings presents challenges, particularly because power plants operate in zones without internet access, while vendors expect to update models, tooling and patches far more often than an isolated zone allows. A critical risk arises when management misjudges the role of AI agents, overemphasising automation at the expense of human oversight. In these environments, judgement calls are vital, such as deciding whether to activate or shut down a switch during an emergency. Human decision-making remains indispensable, as reliance solely on AI could compromise safety and operational integrity.

The OT regulatory burden should move from architectural diagrams toward operational reality. It is key that a control model adequately treats the OT environment as a socio-technical and physical control system rather than primarily as a networked computing environment. Critical-

infrastructure law should regulate the integrity of the physical process, not merely the connectivity of the computers that control it. The ask for OT operators is to demonstrate that they can provide a controlled pathway that prevents a compromised person or an unauthorised device from causing an unsafe or unacceptable physical state.

The appropriate decision-maker will therefore need to weigh operational continuity, safety, statutory obligations and public consequences instead of comparing security expenditure against productivity. OT governance moves beyond a CISO: physical security, engineering, safety, operations, legal and compliance, and executive management all have legitimate responsibilities.

In an OT environment, it is essential to stay within boundaries set by consequence and reversibility. For example, an AI agent might report: "Pump P-6 displayed inconsistent temperature patterns, recommend inspection as soon as possible." Alternatively, the agent could decide on its own to shut down Pump P-6. If the shutdown is a protective function, it belongs to the plant's engineered safety system, designed and tested for that purpose under the applicable functional-safety standard, not to an agent. If it is not a protective function, the unannounced shutdown may cause other pumps to fail. Hence, agents could own observation, correlation and analysis, while humans retain judgement on changes to the physical process and on safety and security exceptions. If an agent is allowed to act autonomously, the authorisation should be explicit, bounded, auditable, tested and revocable, and not based on a general directive.

4. From code signing to evidence

In a regulated environment, non-human identity governance is non-negotiable. Code signing is a critical requirement for validating the authenticity and integrity of AI agents. Without code signing, an organisation cannot show which code its agents were running, which is the first link in any audit trail. Unlike human job roles, where accountability can be directly assigned to individuals, AI agents present unique challenges in establishing responsibility. To address this, AI agents' actions must be made attributable through robust cryptographic mechanisms. Code signing and the related cryptographic controls described below make agent actions traceable, while accountability stays with the people and organisations behind them.

Code signing answers one question well. Provided the signing key is trusted and properly held, it shows that an artefact is the one its publisher released, unchanged since. For an AI agent that is a necessary start. It says nothing, however, about what ran at runtime, what the agent was allowed to do, what it actually did, or who decided to let it act. A supervisor, an auditor or a court will ask these questions after an incident, and the answers have to exist beforehand.

A log can serve as evidence, although its weight depends on how it was generated, protected, retained and attributed. Most logs follow a format a developer chose and sit in a store an administrator can edit or delete. What an organisation can defend before a third party is a

record that is authentic, intact, timed by a trusted source and linked to an accountable person or organisation. In our experience, most agent deployments today produce plenty of telemetry and very few such records.

Extending the code-signing logic across an agent's life means signing four kinds of record. The first, where code signing already sits, is a manifest of what the agent runs: code, model, configuration, policy and tool versions. The second is its authorisation, with the systems it may touch, on whose behalf, for how long, and who approved it. The third is each consequential action, signed as it is taken by the gateway or signing service through which the agent acts, with its inputs and tool calls. A payment instruction or a configuration change then carries its own proof of origin. What counts as consequential is for the organisation's risk model to define. The last is the outcome, captured where possible by the affected system itself, since what an agent reports and what actually changed can differ. The signing keys have to sit outside the agent's reach, or a compromised agent could sign false records about itself. For financial entities, DORA's technical standard already requires controls that protect cryptographic keys against unauthorised access and modification throughout their life (the RTS on ICT risk management, Commission Delegated Regulation (EU) 2024/1774, Article 7(2)) [3].

Trusted timestamps, hash chaining and records of key status let an investigator see that no recorded entry was later altered, reordered or removed. That holds only if the head of the chain is regularly anchored outside the reach of whoever could alter the store. In the EU, a qualified electronic time stamp is presumed accurate as to date, time and data integrity, and a qualified electronic seal as to integrity and origin (eIDAS Regulation, Regulation (EU) No 910/2014, Articles 41(2) and 35(2)) [4]. These presumptions attach where a qualified trust service is used. These controls cannot show that an action went unrecorded, which is one more reason to capture the outcome at the affected system. Together, these records may help reconstruct the chain from approval to effect, which the "AI Superman" of section 2 cannot do from a dashboard.

This design also requires a distinct identity for each agent. Many agents today run on a service account or API key borrowed from a team or application, so every action is attributed to that shared credential. When five agents share one key, the record does not show which agent acted or on whose instruction. An agent-level identity, bound to the signed manifest and the authorisation that launched it, lets a signature point to a specific agent deployment. The agent has no legal personality, and accountability still rests with a natural or legal person, the owner who answers for the agent and can revoke it. This extends the non-human identity governance described at the start of this section, with the agent entered in the control model as a named, owned and revocable component.

This does not have to wait for AI-specific law. In the European Union, at least two instruments already in force reach agents through technology-neutral definitions [1, 2, 5]. DORA, which has applied to financial entities since January 2025, treats software in their systems as an ICT asset. An agent a bank deploys therefore falls within its inventory duty (Article 8), its access and

strong-authentication requirements (Article 9(4)(c) and (d)), and its duties to detect anomalies (Article 10) and to record and report ICT-related incidents, major ones to the supervisor (Articles 17 and 19). Its technical standard points the same way, requiring tamper-protected logs and synchronised clocks (Article 12(2)(d) and (f)), unique identification of natural persons and systems (Article 20) and limits on generic and shared accounts (Article 21) [3].

NIS2 reaches the same ground through its definition of network and information system (Article 6(1)), for essential and important entities outside DORA's scope and where it has been transposed. Its minimum measures include access control and, where appropriate, multi-factor or continuous authentication (Article 21(2)(i) and (j)). For significant incidents, its reporting cascade opens with an early warning within 24 hours of awareness (Article 23(4)(a)). Neither text uses the word agent in the AI sense or prescribes signed agent records. Such records would, however, help an entity show that its access, monitoring and incident duties hold for the agents it runs.

The signatures must also remain verifiable as long as the records are kept. Retention periods vary by sector and jurisdiction and can run to several years when a dispute or investigation reopens a file. Data protection still limits how long personal data may be kept (GDPR, Article 5(1)(e)) [6]. Widely used RSA and elliptic-curve signatures would be vulnerable to a sufficiently capable quantum computer.

NIST finalised its general-purpose post-quantum signature standards in August 2024 (FIPS 204 and FIPS 205) [7, 8]. Its transition guidance, NIST IR 8547 [9], still an initial public draft and not binding, proposes deprecating signature schemes at 112-bit security strength, RSA-2048 for instance, after 2030 and disallowing all quantum-vulnerable ones after 2035. A record signed in 2026 and kept for ten years must still verify in 2036, after both dates. In the United States, Executive Order 14412 (June 2026) [10] directs the Office of Management and Budget (OMB) to require federal agencies to move high value assets and high impact systems to post-quantum digital signatures by 31 December 2031. An organisation that starts signing agent evidence today should therefore choose tooling that can move to post-quantum signatures and plan to re-protect its archive before the old algorithms or keys stop being trusted. In the EU, data preserved by a qualified electronic archiving service enjoy a presumption of integrity and origin for the whole preservation period (eIDAS Regulation, Article 45i(2), as inserted by Regulation (EU) 2024/1183) [4]. DORA's technical standard already asks financial entities to provide for updating their cryptography as cryptanalysis develops (Article 6(4)) [3], although it sets no post-quantum date.

5. The ratio question: two paths and their consequences

Path A (Alan). This AI overdrive has created a blind spot, where AI agents could face internal destruction from rogue agents. Artificial intelligence lacks the intrinsic cognitive reasoning processes that guide human decision-making. Without proper oversight and governance, an AI agent designed for financial tasks could abruptly shift to performing unrelated functions, such as

network testing. Likewise, in the absence of control mechanisms, one AI agent may delegate responsibilities to another, creating unintended chains of action. Hence, the security management requirement for AI agents should preferably be one human to one AI agent.

Path B (Amin). Path B starts from the risk Alan describes and draws a different conclusion. An agent that drifts from its task, or quietly hands work to another agent, is dangerous because nobody can see or bound what it does. Adding people does not fix that by itself. A person assigned to a single agent still cannot tell from telemetry alone which instruction it followed, which child agent it called or what it changed.

Give each agent a distinct identity, a time-limited authorisation and a protected record of its consequential actions. A finance agent that starts probing the network is then acting outside its signed scope, and the action can be blocked or flagged automatically. That requires the authorisation to be enforced as well as recorded, at the gateway through which the agent reaches tools, APIs and the network, which refuses calls outside the signed scope and records the refusal. A delegation to another agent has to carry its own, narrower authorisation, which keeps the chain visible.

One person can then oversee several agents, since the controls do the routine watching and the person handles the exceptions. On this path the ratio depends on factors such as the consequence and reversibility of what the agents do, as the Pump P-6 example in section 3 shows, and a single high-consequence agent may still need a dedicated human. For high-risk AI systems the AI Act (Regulation (EU) 2024/1689) [11] will also bear on the ratio on either path, from the dates in section 6, since providers must design high-risk systems for effective human oversight (Article 14) and deployers must assign that oversight to natural persons with the necessary competence, training and authority (Article 26(2)).

Consequences. Both paths keep people accountable. Path A relies on dedicated human supervision, which keeps a person close to every agent for as long as the organisation funds it. Path B relies on enforceable authorisations and auditable records, which organisations tend to fund once a supervisor or an auditor asks, and existing law already lets them ask (section 4). In practice most organisations will combine the two.

Where regulation is loose, the incentive runs the other way. Organisations are then tempted to let agents take on as much work as they can and to clear the human workforce away, and, in Alan's view, path A is what keeps people in the loop.

Data protection law shows how differently the pressure can be set. If an autonomous AI agent's mistake leads to a personal data breach, Singapore's Personal Data Protection Act (PDPA) [12] allows a financial penalty of up to 10% of the organisation's annual turnover in Singapore where that turnover exceeds S\$10 million, and up to S\$1 million otherwise. The EU's General Data Protection Regulation (Regulation (EU) 2016/679, GDPR) puts a failure of the security duty (Article 32) in the €10 million or 2% tier (Article 83(4)), and reaches €20 million or 4% of worldwide annual turnover, whichever is higher, only where the breach also infringes a basic

principle such as integrity and confidentiality (Articles 5(1)(f) and 83(5)). Singapore's percentage is higher, 10% against 4% at most, but its turnover base is far narrower: turnover in Singapore against worldwide turnover. Enforcement goes beyond fines on both sides. GDPR authorities can restrict or ban processing, and Singapore's regulator, which could already issue directions and had accepted voluntary undertakings under its 2019 enforcement guide [13, 14], has held a statutory power to accept undertakings (section 48L) since 1 February 2021, while the higher penalty cap has applied since October 2022.

If path B prevails, the people who stay need different skills: setting and testing authorisations, reading agent records, and handling the exceptions the controls escalate. Training for those roles is part of the evidence layer, not an afterthought.

6. Three regions, three speeds

Southeast Asia. Notably, there are clear incentives for organisations to replace human labour with AI agents. Unlike human employees, AI agents do not require contributions to pension funds, entitlement to annual or medical leave, bonuses or attention to emotional well-being. This cost efficiency is particularly appealing to small and medium-sized enterprises (SMEs), as it directly supports headcount reduction and is often framed as “business transformation” through the adoption of AI-powered tools to enhance productivity. In contrast, multinational corporations (MNCs) with larger IT budgets typically rely on established vendors such as Google or Microsoft to manage and deploy their AI solutions.

The result is an authority mismatch. The Chief Information Officer (CIO) controls much of the technology budget and operational machinery, whereas the CISO ends up accountable for risks they cannot materially change. CIO responsibility expands horizontally across the enterprise while CISO authority remains vertically inside IT. A CISO with neither budget nor operational authority has responsibility without control, which would also mean nominal security authority over AI agents.

IT is treated as the operational infrastructure, cybersecurity as protection around it, and data governance as an additional layer of discipline. Cybersecurity and data governance are being framed as properties of IT, a legacy organisational hierarchy Alan has observed in Southeast Asia.

In Singapore, Minister for Digital Development and Information Josephine Teo said at the Singapore International Cyber Week in October 2025 that few organisations had embarked on quantum-safe migration, likely because of uncertainty over quantum developments and the lack of specific guidance [15], and the Cyber Security Agency (CSA) released a Quantum-Safe Handbook and Quantum Readiness Index for consultation the same week [16]. Version 1 of the Handbook (July 2026) [17] now lists dated milestones that CSA presents as requirements for critical information infrastructure owners, although the Handbook itself is guidance rather than a binding instrument: a migration plan submitted to CSA by 31 March 2027, newly procured

systems that support quantum-safe algorithms or are quantum-safe ready from 1 January 2028, and migration completed by 31 December 2031. The Monetary Authority of Singapore (MAS) and Banque de France had already run a quantum-safe email experiment in 2024 [18], using CRYSTALS-Kyber and CRYSTALS-Dilithium, since standardised as ML-KEM and ML-DSA [19, 7]. MAS also subjects a broad range of financial institutions to its Technology Risk Management Notices (for banks, MAS Notice FSM-N05 under the Financial Services and Markets Act 2022, which replaced Notice 644 on 10 May 2024) [20, 21].

Singapore has also started on agents themselves. IMDA's Model AI Governance Framework for Agentic AI (January 2026, updated May 2026) [22] recommends that each agent have its own identity, tied to a human user, a supervising agent or an organisational department for accountability. Between November 2025 and January 2026, MAS consulted on guidelines on AI risk management that cover AI agents [23]. It issued them on 7 October 2026, with effect from 7 October 2027 and the life cycle controls expected by 7 October 2028 [24].

The Gulf and Jordan. Field work across the region over the past decade points to a regulator-led pattern, where organisations move once a supervisor sets a date and then tend to move fast. In post-quantum cryptography, the UAE introduced dedicated post-quantum controls in its national information assurance standard in 2025 [25], which applies to federal ministries and authorities and to non-government critical information infrastructure entities. Saudi Arabia's national cryptographic standard, in its 2020 edition [26], is built "assuming the absence of quantum computing capabilities" and confines post-quantum cryptography to a short appendix that names no algorithm and defers the subject to later editions. A revised edition went through public consultation in December 2025 [27] and, as of early October 2026, had not been published [28]. Outside the Gulf, Jordan's central bank launched in January 2026 a roadmap for moving its financial and banking sectors to quantum-resistant cryptography [29]. On AI agents, several executives one of us met in the Gulf in summer 2026 asked who answers for an agent that executes transactions, and none reported a clear answer from vendors or insurers. If agent evidence follows the same path as post-quantum cryptography, it is likely to arrive as an expectation set by supervisors. This is an inference from how the region has handled post-quantum cryptography, not from any rule on agents.

The EU. In Europe the push comes from instruments already in force, since the obligations in section 4 apply through DORA and, where transposed, NIS2. Supervisors have started to read them against AI, so far mainly as a threat. On 7 July 2026 the European Central Bank asked every significant institution it supervises for an action plan on AI-enabled cyber threats by 31 October 2026, adding that post-quantum adoption "must start now" and announcing a separate letter on the quantum risk to encryption [30]. The European Systemic Risk Board published a warning on systemic cyber risks from frontier AI models the same day [31]. The coordinated post-quantum roadmap published by the NIS Cooperation Group in June 2025, which is not binding, asks Member States to start the transition by the end of 2026 and to protect high-risk use cases by the end of 2030 [32]. AI-specific rules then arrive on top, since the Digital Omnibus on AI (Regulation (EU) 2026/1744, in force since July 2026) [33] moved the high-risk obligations to 2

December 2027 for Annex III systems and to 2 August 2028 for systems covered by Annex I product legislation. Once those obligations apply, high-risk AI systems must allow automatic logging of events (AI Act, Article 12), and providers and deployers must keep the logs for at least six months (Articles 19 and 26(6)). Of the regions covered here, the EU is the one where binding ICT-risk duties that reach agents already apply across the eighteen NIS2 sectors and the financial entities within DORA's scope, rather than, as in Singapore, across critical information infrastructure and the financial sector, although their application differs between Member States. On 8 July 2026 the Commission referred Ireland, Spain, France and the Netherlands to the Court of Justice for failing to notify full transposition [34]. The Dutch law, adopted on 7 July but not yet notified when the referral was decided [34], entered into force on 15 August 2026 [35], and as of early October 2026 public trackers still showed France, Ireland and Spain without a NIS2 law in force [36, 37, 38]. The French bill had been scheduled for the Assemblée nationale floor from 7 October 2026, but on 6 October the Assembly's Conference of Presidents left it off an agenda that runs to mid-November [39, 40, 41].

7. Conclusion

An accountability arrangement that exists only on an organisation chart fails at the first incident, whether the chart names one person for ten agents or one person per agent. Accountability holds when the organisation can show, after the fact, which agent acted, under whose authority, within which limits and with what effect. That makes the evidence the first question and the staffing ratio the second. Our advice follows from it: build the evidence layer first, then decide how many people supervise how many agents, according to what the agents can change and how reversibly. The regions will get there at different speeds. In Southeast Asia, the next step Alan would name is to strengthen non-human identity governance, with governments taking an active regulatory role over the providers of the models, clouds and platforms on which agents run. In the Gulf and Jordan, supervisors are likely to set the expectation, as they are beginning to for post-quantum cryptography. In the European Union, instruments already in force reach agents today, and organisations that build the records now will have them when supervisors ask.

Sources

1. European Parliament and Council, Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (DORA), 14 December 2022, OJ L 333, 27.12.2022, Articles 3(7), 8, 9(4)(c) and (d), 10, 17, 19 and 64, <https://eur-lex.europa.eu/eli/reg/2022/2554/oj>
2. European Parliament and Council, Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2 Directive), 14 December 2022, OJ L 333, 27.12.2022, Articles 6(1), 21(2)(i) and (j) and 23(4)(a), Annexes I and II (eighteen sectors), <https://eur-lex.europa.eu/eli/dir/2022/2555/oj>
3. European Commission, Commission Delegated Regulation (EU) 2024/1774 supplementing Regulation (EU) 2022/2554 with regard to regulatory technical standards specifying ICT risk management tools, methods,

processes and policies (RTS on ICT risk management), 13 March 2024, OJ L, 25.6.2024, Articles 6(4), 7(2), 12(2)(d) and (f), 20 and 21, point (c), https://eur-lex.europa.eu/eli/reg_del/2024/1774/oj

4. European Parliament and Council, Regulation (EU) No 910/2014 on electronic identification and trust services (eIDAS Regulation), 23 July 2014, as amended by Regulation (EU) 2024/1183 of 11 April 2024, consolidated text of 18 October 2024, Articles 35(2), 41(2) and 45i(2), <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02014R0910-20241018>
5. Amin Hasbini, The Evidence Question: Regulated Agents Under DORA and NIS2 (Series II, paper 3), 15 July 2026, version 2 of 22 September 2026, Zenodo, concept DOI <https://doi.org/10.5281/zenodo.21325134>
6. European Parliament and Council, Regulation (EU) 2016/679 (General Data Protection Regulation), 27 April 2016, OJ L 119, 4.5.2016, Articles 5(1)(e) and (f), 32, 58(2)(f), 83(4) and 83(5), <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
7. National Institute of Standards and Technology (NIST), FIPS 204, Module-Lattice-Based Digital Signature Standard (ML-DSA), 13 August 2024, <https://csrc.nist.gov/pubs/fips/204/final>, accessed 7 October 2026
8. NIST, FIPS 205, Stateless Hash-Based Digital Signature Standard (SLH-DSA), 13 August 2024, <https://csrc.nist.gov/pubs/fips/205/final>, accessed 7 October 2026
9. NIST, IR 8547 (initial public draft), Transition to Post-Quantum Cryptography Standards, 12 November 2024, section 4.1.1 and Table 2, <https://csrc.nist.gov/pubs/ir/8547/ipd>, accessed 7 October 2026
10. The White House, Executive Order 14412, Securing the Nation Against Advanced Cryptographic Attacks, 22 June 2026, 91 FR 38483 (25 June 2026), Sec. 4(b)(iii), <https://www.govinfo.gov/content/pkg/FR-2026-06-25/pdf/2026-12909.pdf>, accessed 7 October 2026
11. European Parliament and Council, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act), 13 June 2024, OJ L, 12.7.2024, Articles 12, 14, 19(1), 26(2) and 26(6), <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
12. Republic of Singapore, Personal Data Protection Act 2012, as amended by the Personal Data Protection (Amendment) Act 2020 (Act 40 of 2020), sections 48I, 48J(3) (in force 1 October 2022) and 48L (in force 1 February 2021), Singapore Statutes Online (current version), <https://sso.agc.gov.sg/Act/PDPA2012>, accessed 7 October 2026
13. Personal Data Protection Commission (PDPC), Guide on Active Enforcement, 22 May 2019, current edition 7 November 2022, <https://www.pdpc.gov.sg/organisations/resources/guidance-by-topic/guide-on-active-enforcement>, accessed 7 October 2026
14. PDPC, Voluntary Undertaking by HSBC Bank (Singapore) Limited, 10 September 2020, <https://www.pdpc.gov.sg/organisations/regulations-decisions/enforcement-decisions/voluntary-undertaking-by-hsbc-bank-singapore-limited>, accessed 7 October 2026
15. Cyber Security Agency of Singapore (CSA), Opening Address by Minister Josephine Teo at HLP (AI), Singapore International Cyber Week, 22 October 2025, paragraph 26, <https://www.csa.gov.sg/news-events/speeches/opening-address-by-minister-josephine-teo-at-hlp-ai-on-22-oct-2025/>, accessed 7 October 2026
16. CSA, CSA Releases A Quantum-Safe Handbook And Quantum Readiness Index (press release, public consultation from 23 October to 31 December 2025), 22 October 2025, <https://www.csa.gov.sg/news-events/press-releases/csa-releases-a-quantum-safe-handbook-and-quantum-readiness-index/>, accessed 7 October 2026
17. CSA, GovTech and IMDA, Quantum-Safe Migration Handbook, Version 1, 16 July 2026, disclaimer (p. 3) and milestones for critical information infrastructure owners (p. 25), <https://isomer-user->

content.by.gov.sg/36/350e4bd5-fa25-4609-8e78-5e159d9da245/CSA%27s%20Quantum-Safe%20Handbook%20V1%20Dtd%2016%20Jul%2026.pdf, accessed 7 October 2026

18. Monetary Authority of Singapore (MAS), Banque de France and Monetary Authority of Singapore Conduct Groundbreaking Post-quantum Cryptography Experiment to Enhance Communication Security (media release), 5 November 2024, <https://www.mas.gov.sg/news/media-releases/2024/bdf-and-mas-conduct-groundbreaking-post-quantum-cryptography-experiment>, accessed 7 October 2026
19. NIST, FIPS 203, Module-Lattice-Based Key-Encapsulation Mechanism Standard (ML-KEM), 13 August 2024, <https://csrc.nist.gov/pubs/fips/203/final>, accessed 7 October 2026
20. MAS, Notice FSM-N05 Technology Risk Management, issued 9 May 2024 under section 29(1) of the Financial Services and Markets Act 2022, in effect from 10 May 2024, <https://www.mas.gov.sg/regulation/notices/notice-fsm-n05>, accessed 7 October 2026
21. MAS, Notice 644 Technology Risk Management [Cancelled], cancelled with effect from 10 May 2024, <https://www.mas.gov.sg/regulation/notices/notice-644>, accessed 7 October 2026
22. Infocomm Media Development Authority (IMDA), Model AI Governance Framework for Agentic AI, Version 1.0, 22 January 2026 (current Version 1.5, 20 May 2026, updated 5 June 2026), section 2.1.2, Agent identity, p. 23, <https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf>, accessed 7 October 2026
23. MAS, Consultation Paper on Proposed Guidelines on Artificial Intelligence Risk Management for Financial Institutions (P017-2025), open from 13 November 2025 to 31 January 2026, <https://www.mas.gov.sg/publications/consultations/2025/consultation-paper-on-guidelines-on-artificial-intelligence-risk-management>, accessed 7 October 2026
24. MAS, Guidelines on Artificial Intelligence Risk Management for Financial Institutions, published 7 October 2026, in effect from 7 October 2027 (paragraph 1.8), <https://www.mas.gov.sg/regulation/guidelines/guidelines-on-artificial-intelligence-risk-management-for-financial-institutions>, accessed 7 October 2026
25. UAE Cyber Security Council, UAE Information Assurance Standard, Version 2.1, November 2025, section 1.2 (scope) and control T3.4.4 (Cryptographic Requirements in Post Quantum), <https://csc.gov.ae/en/w/uae-information-assurance-standard> (not reachable on 7 October 2026, archived copy of 3 April 2026 accessed that day: <https://web.archive.org/web/20260403041121/https://csc.gov.ae/en/w/uae-information-assurance-standard>)
26. National Cybersecurity Authority of Saudi Arabia (NCA), The National Cryptographic Standards (NCS-1:2020), July 2020, section 1.1 and Appendix 7.2, https://cdn.nca.gov.sa/ar/ncs_en.pdf, accessed 7 October 2026
27. Digital Policy Alert, National Cybersecurity Authority closed public consultation on updates to the National Cryptography Standards (NCS-2:2025), 25 December 2025 (consultation opened 10 December 2025), <https://digitalpolicyalert.org/event/36914-national-cybersecurity-authority-closed-public-consultation-on-updates-to-the-national-cryptography-standards-ncs-22025>, accessed 7 October 2026
28. NCA, The National Cryptographic Standards (NCS), page last updated 16 May 2025, still serving NCS-1:2020, <https://nca.gov.sa/en/regulatory-documents/frameworks-and-standard-list/ncs/>, accessed 7 October 2026
29. Jordan News, Central Bank of Jordan Launches Roadmap for Quantum-Resistant Encryption, 4 January 2026, <https://www.jordannews.io/Section-112/Economy/Central-Bank-of-Jordan-Launches-Roadmap-for-Quantum-Resistant-Encryption-47836>, accessed 7 October 2026

30. European Central Bank, Banking Supervision, letter SSM-2026-0301 to significant institutions on addressing AI-enabled cybersecurity threats, 7 July 2026,
https://www.bankingsupervision.europa.eu/press/letterstobanks/shared/pdf/2026/ssm.2026_letter_on_AI_enabled_cybersecurity_threats.en.pdf, accessed 7 October 2026
31. European Systemic Risk Board, Warning ESRB/2026/3 on systemic cyber risks stemming from frontier artificial intelligence models, 25 June 2026, published 7 July 2026, OJ C/2026/3795, 16.7.2026,
https://www.esrb.europa.eu/pub/pdf/warnings/esrb.warning260625_on_systemic_cyber_risks_stemming_from_frontier_ai_models-ef424708cf.en.pdf, accessed 7 October 2026
32. NIS Cooperation Group, A Coordinated Implementation Roadmap for the Transition to Post-Quantum Cryptography, Part 1, version 1.1, dated 11 June 2025, published by the Commission on 23 June 2025,
<https://digital-strategy.ec.europa.eu/en/library/coordinated-implementation-roadmap-transition-post-quantum-cryptography>, accessed 7 October 2026
33. European Parliament and Council, Regulation (EU) 2026/1744 amending Regulation (EU) 2024/1689 (Digital Omnibus on AI), 8 July 2026, OJ L, 24.7.2026, in force 27 July 2026, amended Article 113, third paragraph, point (c)(i) and (ii), <https://eur-lex.europa.eu/eli/reg/2026/1744/oj>
34. European Commission, Commission refers Ireland, Spain, France and the Netherlands to the Court of Justice for failing to transpose the rules on cybersecurity, 8 July 2026, <https://digital-strategy.ec.europa.eu/en/news/commission-refers-ireland-spain-france-and-netherlands-court-justice-failing-transpose-rules>, accessed 7 October 2026
35. Government of the Netherlands (Rijksoverheid), Cyberbeveiligingswet en Wet weerbaarheid kritieke entiteiten vanaf 15 augustus 2026 van kracht, 7 July 2026,
<https://www.rijksoverheid.nl/actueel/nieuws/2026/07/07/cyberbeveiligingswet-en-wet-weerbaarheid-kritieke-entiteiten-vanaf-15-augustus-2026-van-kracht>, accessed 7 October 2026
36. Sildaro, NIS2 en France : la loi n'est toujours pas là, mais l'horloge tourne, 31 August 2026,
<https://sildaro.com/cyber-attaque/nis2-france-transposition-etat-des-lieux-aout-2026/>, accessed 7 October 2026
37. 6clicks, Ireland: NIS2 isn't law yet, but you're already on the hook, 10 August 2026,
<https://www.6clicks.com/resources/blog/ireland-nis2-isnt-law-yet-but-youre-already-on-the-hook>, accessed 7 October 2026
38. nisd2.eu, Estado de NIS 2 en España, last checked 5 October 2026, <https://nisd2.eu/es/wiki/timelines-and-status/nis2-status-spain>, accessed 7 October 2026, and NIS2 tracker: transposition status in all 27 EU countries, as of 4 October 2026, <https://nisd2.eu/en/wiki/timelines-and-status/nis2-eu-tracker>, accessed 7 October 2026
39. Assemblée nationale, Projet de loi relatif à la résilience des infrastructures critiques et au renforcement de la cybersécurité, legislative file,
<https://www.assemblee-nationale.fr/dyn/17/dossiers/DLR5L17N50731>, accessed 7 October 2026
40. Assemblée nationale, Ordre du jour de l'Assemblée nationale (Conférence des présidents du mardi 6 octobre 2026), agenda to 17 November 2026, rolling file,
<https://www.assemblee-nationale.fr/dyn/static/seance/conference.pdf>, consulted 7 October 2026, copy saved
41. Next, [MàJ] Caramba, encore raté : la transposition de NIS2 est de nouveau repoussée, published 22 September 2026, retitled and updated 6 October 2026 (the URL keeps the original slug),
<https://next.ink/brief-article/alleluia-la-transposition-de-nis2-est-enfin-a-lordre-du-jour-de-lassemblee-nationale/>, accessed 7 October 2026

About the authors

Amin Hasbini is an AI and cybersecurity executive based in Paris, working on AI-agent security and post-quantum readiness. He spent twelve years in threat research, the last seven directing Kaspersky's Global Research and Analysis Team (GReAT) for the Middle East, Turkey and Africa.

Alan Seow is a cybersecurity governance, risk and compliance leader based in Singapore. The Southeast Asia observations in this paper come from his field work with organisations in the region.