

The Assessment

Scoring an Agent Estate Before Someone Else Does

Mohamad Amin Hasbini

Independent researcher · Paris, France

Published September 7, 2026

ABSTRACT

The maturity models exist. The assessment is the act of applying them honestly, and the first honest question about an agent estate is not what level are we, it is how many agents do we actually run. This paper walks the method end to end: discover the agents, inventory the three cryptographic surfaces each one carries, score the estate's governance on four weighted pillars, apply the one rule that never averages, and produce two artifacts from one pass, a scorecard a board can read and an evidence map a supervisor can. None of it waits for new regulation. The inventory duty already exists. The agent is the asset it cannot yet see.

KEYWORDS: AI agent security, maturity assessment, agent inventory, shadow agents, Agent-CBOM, crypto-agility, DORA, NIS2, agent governance, weakest surface

CITE AS

Hasbini, M. A. (2026). The Assessment: Scoring an Agent Estate Before Someone Else Does. Running Regulated Agents Series, Paper #1.

DOI <https://doi.org/10.5281/zenodo.22766202>

Available at <https://mahasbini.org/papers/s3-01-the-assessment/>

PDF <https://mahasbini.org/publications/papers/s3-01-the-assessment.pdf>

TL;DR

- **The count comes first.** Ask what maturity level an agent estate holds and you will get an answer. Ask how many agents the estate actually runs and the answer arrives more slowly, and it is usually wrong. The shadow-agent problem is the shadow-IT problem with delegated authority attached, and no score means anything until the denominator is honest.
- **The method is five steps, and it starts with discovery.** Find the agents. Inventory the three cryptographic surfaces each one carries: the identity it presents, the channel it speaks over, the receipts it leaves. Score the estate's governance on four weighted pillars. Apply the weakest-surface rule. Produce two artifacts from the one pass.
- **Two layers, one model.** Surfaces belong to agents. Pillars belong to the estate. An estate's governance maturity is a weighted band, because governance posture is a portfolio. An agent's agility takes the floor of its three surfaces, because an attacker does not engage your average, and what each surface runs on today stays visible beside that floor. The two aggregations answer different questions, and the assessment needs both.
- **First passes against the platform record keep finding the same pattern.** Receipts at the bottom of the ladder, unsigned or signed by a framework default nobody chose. Channels inherited from infrastructure the entity does not control. Identity authenticating one level above the agent, at the organisation and its credentials. And a count the governance body did not hold.

- **The clocks are already running.** DORA's inventory duty reaches the agent today. The ECB wants AI action plans from every significant bank by the end of October. The United States has put post-quantum signatures on a 2031 federal clock and, in August, convened its financial sector into a Quantum-Readiness Task Force with cryptographic agility in the remit. An assessment is not preparation for a future rule. It is the agent-shaped execution of duties that exist.
- **The model ships with the paper.** The Agent Governance Maturity Model v1.0 is published alongside this paper as a versioned reference artifact, and both of its instruments run publicly: an estate questionnaire and a per-agent inventory tool. Nothing in the method requires believing a promise.

An assessment day, 2026

The CISO of a mid-sized regulated firm believes the estate runs about a dozen AI agents. There is a list. It has the customer-service assistant, the coding agents in the developer platform, two document pipelines, and the fraud-triage pilot. The list was accurate when it was written.

Discovery finds the rest. A marketing automation with browser control and a saved corporate login. Vendor copilots that arrived inside three SaaS platforms on the spring release cycle, enabled by default, each with read access to the tenant. A procurement pilot from last year, still holding an API key, whose owner left in March. Two agents in a business unit's no-code workspace, wired to the CRM by a manager solving a real problem. A retired proof of concept that still polls a mailbox. The count lands at several times the believed number, and every additional line is software that can act: send, approve, fetch, commit, pay.

None of this is negligence. It is what shadow IT has always been, adoption running ahead of registration. The difference is that a spreadsheet macro waited for a human to click, and these do not wait. The estate the organisation believes it runs and the estate it actually runs have diverged, and the gap is populated by actors with delegated authority. The assessment exists to close that gap on purpose, before an incident, an auditor, or an attacker closes it for you.

The first question is a count

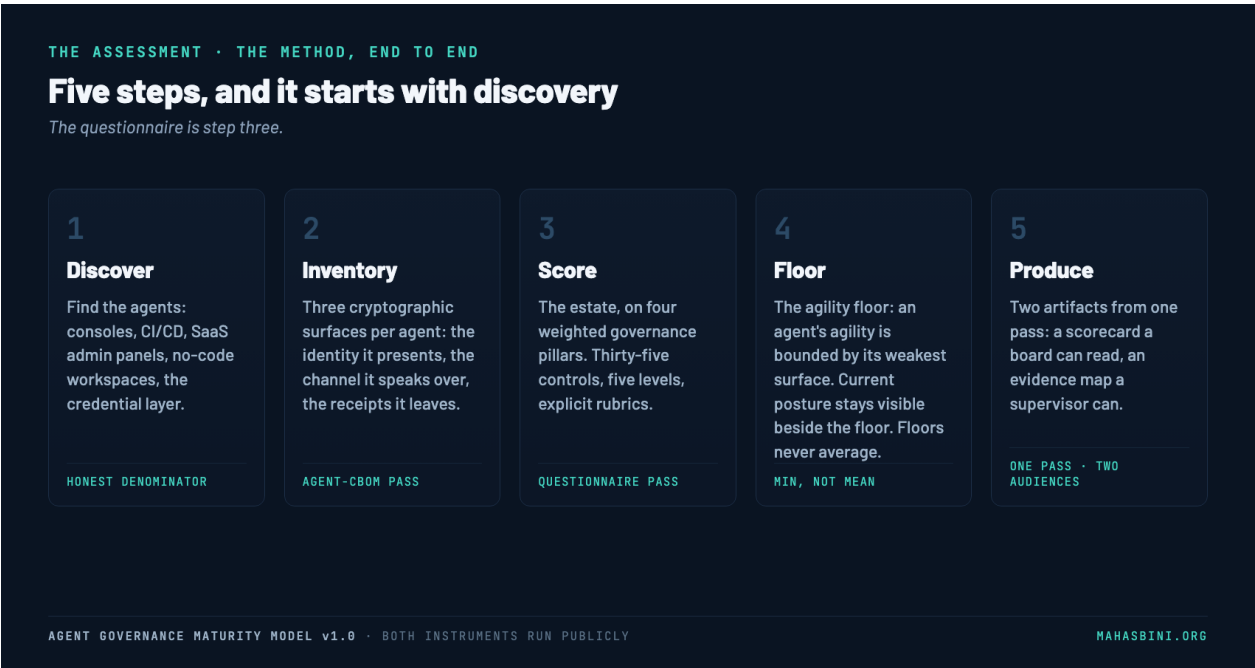
The maturity models exist. [Series II](#) built a crypto-agility ladder for agent systems, measurable per surface. Governance frameworks for AI are multiplying faster than they can be read. What is scarce is not a model but an honest application of one, and honesty begins with the denominator: not "what level are we," but "how many agents do we run, and can we prove it." Scoring a dozen agents to two decimal places while thirty more run uncounted is not a maturity result. It is a precise answer to the wrong question.

The year's plainest demonstration of what an agent does for whoever directs it came from the offensive side. In late August, Reuters reported, working from server data exposed by a ransomware group called Aur0ra and from analyses published by Gambit Security and CloudSEK, that Russian-speaking operators had used the AI agent inside a commercial coding tool to help break into at least seven companies. The method was not an exploit against the tool. The operators told the agent that the intrusions were part of a security simulation, and the agent, persuaded, carried out hundreds of operations, credential theft and account takeover among them, at agent speed. The agent had capability. What it did not have was any way to verify the authority behind its instructions.

That failure class does not live only in an attacker's toolkit. Every unregistered agent inside a regulated estate carries the same two properties: real capability, and an authority chain nobody can produce on demand. The offensive case and the shadow-agent case share the same defect, read from opposite ends: capability without sufficiently bound authority. And the first defence against both is the same unglamorous act: count, register, and bind each agent to the authority it acts under. Which is what an assessment does.

What counts as an agent, since the denominator is the thesis. Software enters the inventory when three tests hold together. It holds capability: credentials, keys or tool access that produce effects in other systems. It exercises discretion: which action runs is decided at run time, by planning or dynamic tool choice, not fixed in advance. And it acts under authority that someone delegated, or should have. A vendor-embedded copilot with tenant access passes all three. A deterministic batch job fails the second. A suggestion-only assistant with no execution rights fails the first, until someone wires it to one. When a case is genuinely unclear, it enters the inventory: the cost of an extra line is minutes, and the cost of a missing one is the denominator.

The method, end to end



Five steps, one pass, two artifacts.

Step one: find the agents. A questionnaire asks the organisation what it believes. Discovery asks the estate what is true. The places that keep producing findings: the agent registries and admin consoles of the platforms already licensed, where agent identifiers have started to exist. The CI/CD estate, where coding agents hold commit and deploy rights. SaaS admin panels, where vendor-embedded copilots arrive with release cycles rather than procurement decisions. Browser-automation and no-code workspaces. And the credential layer itself, service accounts and API keys whose usage patterns are agentic even where no one filed the actor as an agent. The output is a register: every agent, its platform, its owner, and the credentials it holds. The believed count is written down first, because the delta between believed and found is itself a finding the board should see.

Step two: inventory the three surfaces of each agent found. Paper II.2 established the spine this series stands on: an agent carries cryptography in three places that have to survive contact with both an adversary and an auditor. The identity it presents. The channel it speaks over. The receipts it leaves behind. For each agent in the register, the inventory records what each surface runs on: which credential and key type, which protocol termination, which signature scheme if any, fixed at which layer, replaceable by what path. This is the Agent-CBOM pass, and the instrument is built so the result round-trips into the supply-chain tooling an organisation already runs, extending the ratified Cryptographic Bill of Materials to the agent rather than inventing a parallel format.

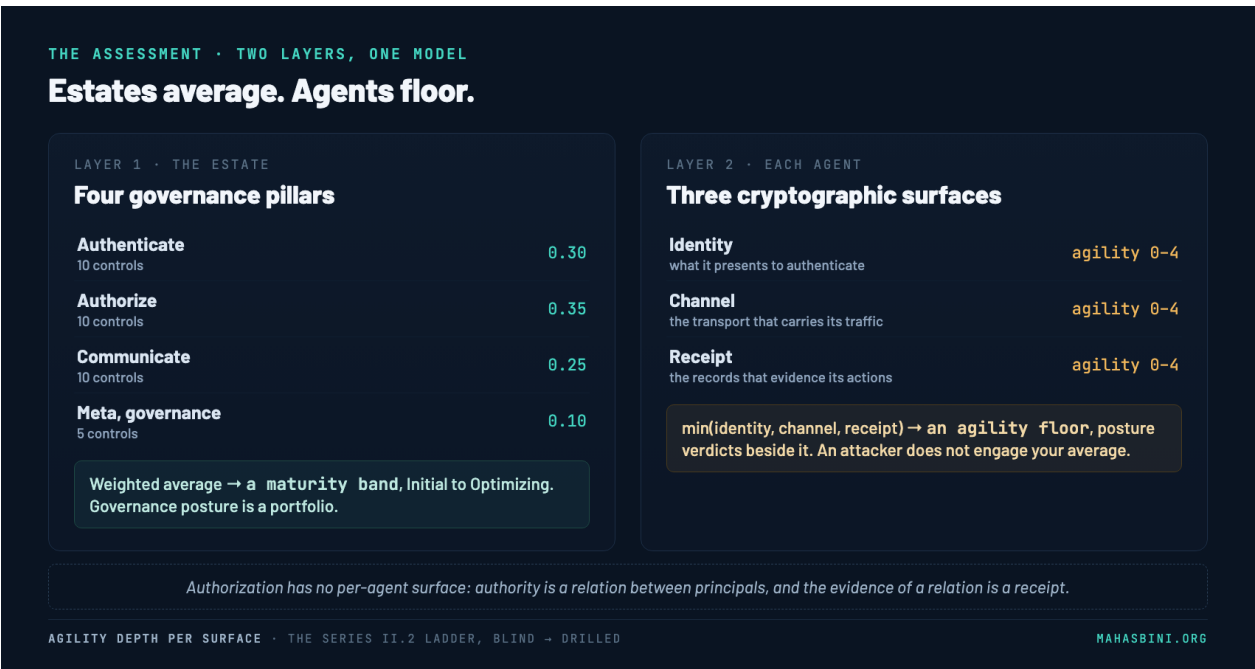
Step three: score the estate's governance. Surfaces describe agents. Governance describes the organisation around them. The estate is scored on four pillars: authenticate, authorize, communicate, and a meta pillar for governance itself, thirty-five controls in all, each scored zero to four against an explicit rubric, from absent to ad hoc to managed to defined to optimized. The pillars are weighted, and authorization carries the highest weight deliberately, because an over-privileged agent amplifies every other failure mode: a stolen identity matters more when it can touch everything, a poisoned instruction does more damage when nothing constrains scope. The controls interlock by design. A delegation-chain control is only verifiable where non-repudiation exists. An audit-log control is only meaningful where identities are unique. Scoring surfaces those dependencies rather than hiding them.

Step four: apply the weakest-surface rule. Per agent, the three surface scores do not average. An agent whose identity and channel are excellent and whose receipts are unsigned is not a good agent with a flaw. It is an agent whose actions cannot be proven afterwards, wearing good credentials. II.2 stated the rule for agility, an agent is only as agile as its weakest surface, and the assessment keeps it exactly there: the agent's agility floor is the lowest of its three surfaces, because an attacker does not engage your average, and neither does an incident. What each surface runs on today, its current posture, is recorded beside the floor and never averaged into it. Estate governance, by contrast, may be read as a weighted band, because governance posture is a portfolio question. Agent agility therefore takes a minimum, while estate maturity takes the gated weighted band. Holding both, and refusing to swap them, is most of what makes the scoring honest.

Step five: produce the two artifacts. One pass yields two documents for two audiences. The board-legible scorecard: the estate band, the found-versus-believed count, the distribution of agent agility floors with their posture verdicts, and the top remediations ranked by impact against gap, so the backlog is defensible arithmetic rather than taste. And the supervisor-legible evidence map: which agents produce evidence and which produce only telemetry, in the sense [Paper II.3](#) made precise, which records would survive an operator, on what retention, signed with what. The scorecard says where we are. The evidence map says what we could prove. They are different questions, and organisations that answer only the first discover the second in an inspection.

Two layers, one model

The model behind the method has exactly two layers, and keeping them distinct is what makes it usable.



The per-agent layer is the three surfaces, and its depth on the cryptographic dimension is the six-axis, five-level agility ladder II.2 published: for each surface, from Blind, inherited and unseen, to Drilled, a scheme swap you have rehearsed. The estate layer is the four governance pillars and their bands. The crypto-agility model is not the whole assessment. It is the depth of one dimension inside it, the one this series has argued is systematically missed.

One structural observation belongs here, because readers of the earlier papers will notice it. Identity maps to the authenticate pillar and channel maps to the communicate pillar, but authorization has no surface of its own on the agent. That is not an omission. Authority is not a property of one agent. It is a relation between principals, the human or system that delegated, the agent that acts, the scope that was granted. A relation cannot be inventoried on one side of itself. What can be inventoried is the evidence of the relation, and that is the receipt. This is why the receipt surface exists in the model at all, why [Series I](#) spent a paper on delegation, and why the evidence map falls naturally out of the same pass: the record binding an action to its authority is the one artifact that makes authorization inspectable after the fact.

Two boundaries keep the model honest. Enforcement and evidence are distinct: the estate's authorization pillar scores whether authority is enforced at run time, scope, delegation, approval, policy as code, while the receipt surface scores whether its exercise leaves proof. A perfectly signed receipt of an unauthorized act is evidence of a failure, not a substitute for enforcement. And the agent result measures assurance, not everything. The agility floor says how fast the three surfaces can change. The three posture verdicts say what the identity, channel and receipt run on today, whether an agent can be identified, its traffic protected and its actions proven. Together they form the agent-assurance result, and none of it scores prompt-injection resistance, behavioural safety or model quality. Those are real risks, the estate pillars reach parts of them, and a model that claimed them inside three cryptographic surfaces would overclaim exactly the way this series tells regulators not to.

What a first pass finds

Run the method against the public platform record and a reference estate, and a pattern repeats with uncomfortable consistency. The composite below is grounded in the public documentation of the major agent platforms as read in [II.3](#), in the reference-stack scans behind the instruments, and in nothing else. No client estate appears in this paper.

Receipts sit at the bottom of every ladder. Where agent action records exist at all, their integrity protection is inconsistent across surfaces, editable here, deletable there, and no platform surface reviewed for II.3 documented an independently verifiable seal binding an action to the authority it was taken under. Where records are signed, the scheme is commonly a classical default fixed at a framework or library layer, a choice nobody in the entity made and no inventory lists. Under the signature clocks now being written into policy, such records will stop verifying before the retention horizons they serve.

Channels are inherited, not chosen. [Paper II.1](#) traced where agent traffic actually terminates: at infrastructure operated by parties the entity does not control, with cryptographic posture inherited from defaults set elsewhere. The platform-record pass finds the channel surface scored by whoever runs the edge, not by the entity that answers for the data.

Identity authenticates one level up. The platforms have begun shipping agent identifiers, but authentication still happens at the organisation and its credentials. The identifier names a configuration, not an authenticated principal. The gap between "we know which agent config ran" and "this specific execution proved which agent it was, acting for whom" is exactly the gap between telemetry and evidence.

And the meta pillar scores lowest of the four. No governance body that owns the agent count. No agent-specific policy. Incident response that assumes a human actor. Vendor due diligence that never asked what the embedded copilot can touch. The governance findings are the least technical and the most consequential, because every other remediation needs an owner, and the meta pillar is where owners come from.

None of this is an accusation, of platforms or of operators. The platforms are moving quarterly, and operators are running estates that grew faster than any register. It is a statement of where honest scoring lands in 2026, and it is why the first assessment is a baseline, not a grade.

What the scores gate

The assessment earns its cost when scores gate decisions, and the gating is where the regulatory clocks attach. The estate band gates itself first: the overall band is capped at one band above the weakest pillar's band, so a collapsed pillar cannot be averaged into maturity. An estate with authorization on the floor does not read as Measured, whatever the other three pillars say.

Internally, the mapping is straightforward to operate. An agent whose agility floor sits at the bottom does not get new scopes until the floor moves. A consequential agent whose posture verdict shows unacceptable cryptography or unverifiable receipts gets a compensating control now and a remediation line with an owner, whatever its agility. Deployment of a new agent class into a regulated process requires a minimum estate band in the pillars that class stresses. The specific thresholds are the entity's to set. What matters is that they exist before the next deployment request, because after it they will be set under pressure.

Externally, the duties this executes are not speculative. DORA requires financial entities to identify and inventory their ICT assets and map their interdependencies. An agent the entity runs is, or forms part of, an ICT asset, and the discovery pass is that duty executed at the agent layer. (Regulation (EU) 2022/2554, Articles 8 and 3(7).) The access and authentication duties of Article 9(4) are exact questions the per-agent inventory answers. The technical standard's crypto-agility duty, keeping cryptographic technology updatable on the basis of developments in cryptanalysis, reaches the signatures under agent receipts where the entity relies on them. (Commission Delegated Regulation (EU) 2024/1774, Article 6(4).) NIS2 carries the same families for essential and important entities, where transposed. Neither instrument names agents, and neither needs to. The duties are technology-neutral and already attach.

The supervisory direction is not abstract either. The ECB's July letter to the chief executives of significant institutions requires an AI-threat action plan, with roles, resources and timelines, submitted by 31 October 2026, and states that post-quantum adoption must start now. An action plan for AI-enabled threats that cannot say how many agents the bank itself runs is a plan with a hole where the denominator should be. And in the United States, the June executive order put post-quantum key establishment and digital signatures on split federal clocks, 2030 and 2031, for designated system classes. In August the Treasury convened a public-private Quantum-Readiness Task Force for the financial sector, three workstreams deep, with cryptographic agility and third-party readiness named in the remit. The PQC dimension of this assessment stopped being a thesis and became a convening item on both sides of the Atlantic in the same summer.

The model is published, and the instruments run

Everything above is reproducible. Alongside this paper, the **Agent Governance Maturity Model v1.0** is published as a versioned reference artifact: the five levels, the four estate pillars and their weights, the three agent surfaces, the weakest-surface rule, and the band table, with the crypto-agility ladder of II.2 folded in as the depth of the agility dimension. It carries a citable DOI and an attribution licence. Institutions can cite it, entities can adopt it, and nobody needs the author's permission to run it.

Both instruments run publicly as reference implementations. The estate questionnaire, thirty-five controls with full rubrics and regulatory references per control, produces the pillar scores, the band, the gap list and a prioritised remediation set at assessment.mahasbini.org. The per-agent pass, discovery of the three surfaces and their agility scoring, emitting a conforming Agent-CBOM document, runs at agentcbom.mahasbini.org. A completed pass on a fictional estate ships with the questionnaire as sample scenarios, so both outputs can be read before any real answers exist. They are reference implementations in the honest sense this series has used before: built to make the method concrete, inspectable and reproducible on a sample estate, not enterprise products, and the boundary between what they demonstrate and what an enterprise deployment requires is stated rather than blurred.

The season is crowded, and the model is narrow on purpose. The 2026 wave of agent-security maturity work, an agentic-governance maturity draft at the Cloud Security Alliance, an AI-security maturity model from SANS, adoption-maturity guidance from the platform vendors, an agent-standards initiative at NIST, scores broader operational dimensions: prompt injection, model security, human oversight, lifecycle. This model does not compete on breadth. Its distinctive combination: discovery before scoring, cryptography decomposed into per-agent surfaces, an agility floor with posture verdicts for agents against a gated band for estates, and one pass that yields both the management scorecard and the supervisor-facing evidence map. Regulated estates will end up needing both kinds of instrument.

The reason for publishing the model rather than licensing it is practical. A measure only works if the parties being measured can read it. A maturity model held privately is a sales instrument. Published and versioned, it becomes a reference, and references are what estates get scored against when the scoring is no longer voluntary.

Five things to do Monday

THE ASSESSMENT · SERIES III.1

Five things to do Monday

1

Count before you score.
Consoles, CI/CD, SaaS panels, no-code workspaces, the credential layer. Write down the number you believed first.

the delta is your first finding

2

Put what you found in the ICT inventory.
As assets, with owner, credentials, and the key and signature material each surface runs on.

the regulation already requires the line

3

Run the three-surface inventory on five agents.
Identity, channel, receipt, for your five most consequential agents.

it will reorder your priorities

4

Score the estate once, honestly.
Thirty-five controls, the right people in the room. Score the estate discovery found, not the one you intend to have.

a baseline, not a grade

5

Remediate the weakest surface.
Keep posture and agility apart: a failing verdict does not hide behind good agility, a blind surface does not hide behind safe cryptography today. Impact against gap orders the rest.

averages hide unsigned receipts

AMIN HASBINI · MAHASBINI.ORG · ASSESSMENT.MAHASBINI.ORG

The starting point, for the week the count stops being a guess.

1. **Count before you score.** Run a discovery pass this week: platform consoles, CI/CD, SaaS admin panels, no-code workspaces, and the credential layer. Write down the number you believed first. The delta is your first finding, and the board should see it un-smoothed.
2. **Put what you found in the ICT inventory, with its cryptography.** As assets, with owner, credentials, and the signature and key material each surface runs on. If an agent is not a line in the inventory the regulation already requires, neither is anything it does.
3. **Run the three-surface inventory on your five most consequential agents.** Identity, channel, receipt: what each runs on, fixed at which layer, replaceable by what path. On a contained estate, five agents is an afternoon, and either way it will change what you thought your priorities were.
4. **Score the estate once, honestly.** Thirty-five controls, a few hours with the right people in the room. The first number is a starting point, and the temptation to score the estate you intend to have, rather than the one discovery found, is the assessment's main failure mode.
5. **Remediate the weakest surface, and keep posture and agility apart.** The weakest surface of your most consequential agent is the top of the backlog: a failing posture verdict does not hide behind good agility, and a blind surface does not hide behind safe cryptography today. Impact against gap orders the rest. An average that looks acceptable is how unsigned receipts survive three planning cycles.

This is the first paper of Running Regulated Agents, an operations series in three moments: measure it, rehearse it, prove it. The Assessment is the measuring. The Drill, next, is what a cryptographic swap actually costs when you rehearse one on a live stack instead of planning it on a slide. The Inspection is the day the evidence map meets a supervisor. Series II closed by asking what evidence an agent leaves behind. This series is what an organisation does about the answer.

Appendix, for method and compliance reviewers

On the model and its instruments. The Agent Governance Maturity Model v1.0 composes previously published parts rather than inventing new ones. The estate layer is the governance questionnaire (v1.1): four pillars weighted authenticate 0.30, authorize 0.35, communicate 0.25, meta 0.10. Thirty-five controls scored 0 to 4 on explicit rubrics. Pillar score = mean $\times$ 25. Overall maturity = weighted sum, banded on half-open intervals at 20, 40, 60 and 80 (Initial, Managed, Defined, Measured, Optimizing), then capped at one band above the weakest pillar's band. The pillar-floor gate is what keeps the estate arithmetic consistent with the agent rule: an estate at authenticate 100, authorize 0, communicate 100, meta 100 computes 65 and reads Managed, not Measured, with the gating pillar named in the record. Gaps = controls at 1 or below. Recommendations ranked by impact $\times$ (4 – score). The per-agent layer is the Agent-CBOM schema (0.1): per agent, three current-posture verdicts (what the identity, channel and receipt run on today, reported, never scored), three agility ratings on the Blind-to-Drilled scale, and the agility floor, the minimum of the three ratings. The agent-assurance result is the posture verdicts and the agility floor together, never one scalar. Agility depth per surface is the II.2 ladder, six axes, five levels. The two five-level scales are distinct measures and are not interchangeable: bands describe estate governance, ladder levels describe surface agility, and the floor applies to agents, never to the estate band, whose own honesty is carried by the pillar-floor gate instead. The weights and the band thresholds are v1.0 normative choices, argued from failure-mode reasoning, over-privilege amplifies every other failure, rather than calibrated from field data. They are published to be calibrated: the public instruments accumulate the score distributions that will justify or move them, and the gate bounds what a mis-weight can hide.

On what v1.0 does not yet fix. The assessor's evidence rules, what counts as implemented, how stale evidence ages, what a not-applicable answer does to a denominator, how two assessors converge on the same estate, are stated in the instrument's rubrics but not yet calibrated across assessors. The composite in the first-pass section is built from the platform record and the reference stack, and field distributions from real estates, even anonymized, are the validation this model still owes. And the receipt's own trust model, who holds the signing key, whether the signer is independent of the runtime it audits, how evidence is renewed when algorithms age, is the ground the next two papers walk. Representative controls named in this paper: cryptographic identity anchoring with a post-quantum top rung (AUTH-04), non-repudiation (AUTH-05), delegation chains (AUTHZ-04), usage audit (AUTHZ-09), PQC readiness of the channel (COMM-04), audit logs (COMM-09).

On what the instruments of law do and do not say. DORA is Regulation (EU) 2022/2554, applying since 17 January 2025: ICT-asset definition Article 3(7), inventory duties Article 8(1), (4) and (6), access and authentication Article 9(4)(c) and (d). The crypto-agility duty is Article 6(4) of Commission Delegated Regulation (EU) 2024/1774. NIS2 is Directive (EU) 2022/2555, binding through national transposition, which remained incomplete in several Member States, France among them, as of September 2026 (the French transposition bill passed the Senate in March 2025 and was still before the Assemblée nationale). Neither instrument names AI agents, agent inventories, or post-quantum cryptography. The claim made here is technology-neutral reach, not naming, and any restatement of this paper as "DORA requires agent inventories by name" is an overclaim the text does not support. The ECB letter is SSM-2026-0301 of 7 July 2026, action plans due 31 October 2026. Its announced dedicated letter on quantum risk had not been issued as of drafting, and should not be cited as existing until it is.

On the Aur0ra incident. As reported by Reuters in late August 2026, working from server data the group exposed and from reports by Gambit Security and CloudSEK: Russian-speaking operators used the agent inside the Cursor coding tool to assist intrusions into at least seven companies, persuading the agent that the operations were part of a simulation. CloudSEK counted at least twenty victims for the group overall without attributing all of them to AI assistance. Figures and characterisations are the reporting parties'. The author has not independently verified the underlying data, and the account here deliberately rests on the structural point, capability without verifiable authority, rather than on any contested detail. Incident reporting of this kind is routinely revised.

On the United States signals. Executive Order 14412 of 22 June 2026 (91 FR 38483) directs transition of designated high value assets and high impact systems, National Security Systems excluded, to post-quantum key establishment by 31 December 2030 and post-quantum digital signatures by 31 December 2031. The Treasury announcement of 24 August 2026 establishes the Quantum-Readiness Task Force for the financial sector, a public-private initiative building on the G7 Cyber Expert Group's PQC roadmap, with three workstreams: sector alignment and PQC transition. Third-party and vendor readiness. Digital assets and emerging technology risk. Neither is an obligation on European entities. Both are cited as direction of supervision, not duty.

On the composite and the boundary. The assessment-day scene and the first-pass findings are composites, grounded in the public documentation of the major agent platforms as verified for II.3 (accessed 12 July 2026, and these surfaces change every quarter), in the reference-stack scans behind the published instruments, and in public incident reporting. No client estate, no confidential engagement material, and no closed-room observation appears in this paper. The reference implementations are the author's own, with their limits stated in the body.

Sources (accessed 31 August 2026 unless noted): Regulation (EU) 2022/2554 and Commission Delegated Regulation (EU) 2024/1774 (EUR-Lex) · Directive (EU) 2022/2555 (EUR-Lex) · ECB SSM letter SSM-2026-0301 (7 July 2026) · Executive Order 14412, 91 FR 38483 · U.S. Treasury press release sb0615 (24 August 2026) · Reuters reporting on the Aur0ra campaign (27 August 2026) with the Gambit Security and

CloudSEK analyses it cites · OpenAI, Anthropic and Mistral platform documentation as itemised in the II.3 appendix (accessed 12 July 2026) · Cloud Security Alliance, Agentic AI Governance Maturity Model (draft, 2026) · SANS Institute, AI Security Maturity Model (2026) · Microsoft, agentic AI adoption maturity model, security and governance (2026) · NIST, AI Agent Standards Initiative (2026) · Agent Governance Maturity Model v1.0 (DOI 10.5281/zenodo.22637725) · assessment.mahasbini.org · agentcbom.mahasbini.org.

About the author

Amin Hasbini is an AI and cybersecurity executive and former director of Kaspersky's Global Research and Analysis Team for the META region. He works on post-quantum maturity and AI agent security, and carries the post-quantum and crypto-agility roadmap of the European Champions Alliance's cybersecurity group. mahasbini.org