daiseQ

The Assurance Paradox: why compliance doesn't equal data protection

On the widening gap between audit-ready compliance and the sensitive data actually exposed, and what senior leaders can do about it.

Brian Wawengkang | Founder, DAISEQ

v1.0 | May 2026

Executive summary

Most well-run organisations are able to pass their security audits. Far fewer can answer, with any real conviction, the question their boards are increasingly asking, *'if we have a cyber breach tomorrow, how quickly can we isolate the blast radius, what is our immediate financial and regulatory exposure on the compromised sensitive data, and how fast can we resume core business operations?'*.

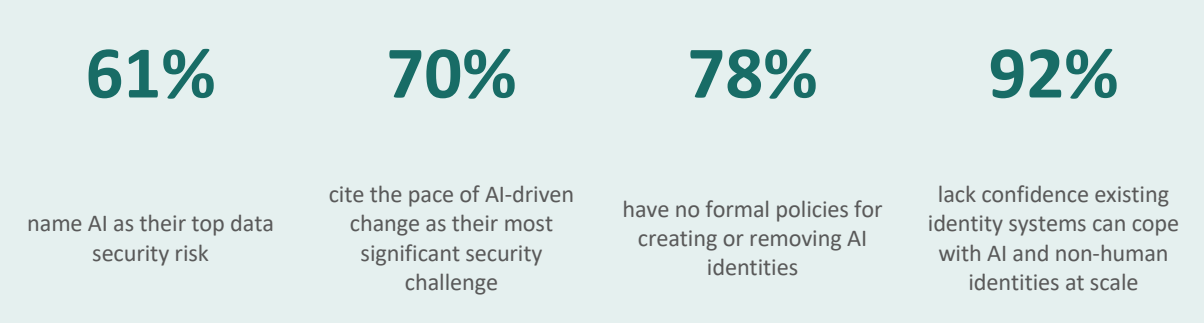
> **What we mean by 'sensitive data'**
>
> For the purposes of this paper, sensitive data is any data whose loss, exposure or misuse would create material financial, regulatory, operational or reputational consequence for the organisation. This includes regulated data (PII, PHI and payment data), commercially sensitive information (M&A activity, intellectual property, pricing and trade secrets) and operationally critical assets (cryptographic keys, production systems, model weights and embeddings). This is, in practical terms, the organisation's Crown Jewels: any information whose mishandling would require escalation to senior leadership or the board.

Sensitive data is the last line of defence. By the time exposure reaches the data itself, every other control has already failed or been bypassed. The gap between being audit-compliant and whether that data is actually exposed has been growing for some years, and with the adoption of AI it is now widening sharply. Sensitive data is more distributed across cloud, SaaS and partner platforms than at any point in the last decade; it is reused more aggressively, by more systems and more identities; and a large proportion of that reuse now happens at machine speed, often without human intervention. AI did not create this dynamic, however, it did compress the timescale on which this plays out, and it has done so exponentially faster than most security assurance controls were designed to handle.

For senior leaders, the defensible position in 2026 is no longer about whether the right controls exist. It is about whether those controls work, end to end, on the data paths that truly matter to the business. That is a different kind of question, and it requires a shift most organisations have yet to make: from static, policy-led assurance… toward continuous, context-aware visibility of how sensitive data actually moves through the estate.

The shift required is also being complicated by the regulatory wave now landing. DORA is already in force across financial services, the EU AI Act is being phased in through 2026 and 2027 across industries, and sector-specific guidance, including from the FDA, EMA and MHRA, is steadily tightening expectations in healthcare and life sciences. Collectively, this is increasing the demand for demonstrable evidence of control effectiveness and accountability in how data and AI systems are governed. Each new regulation will, in my view, also create a fresh wave of false assurance, where organisations mistake regulatory readiness for operational protection. The gap I am describing in this paper is in many cases widened by regulation, not closed by it.

The signal from the market is unusually consistent.



Source: Thales 2026 Data Threat Report; Cloud Security Alliance and Oasis Security 2026. Statistics verified against original published reports as of May 2026.

Governance, in other words, is losing ground to operational reality. The gap between the two is where exposure hides, and the rest of this paper is about how to close it.

1. The Assurance Paradox

Audit success is evidence of process maturity. It is not, on its own, evidence of real-world protection, and conflating the two has become one of the more expensive habits in enterprise security.

In my own work leading global data and cloud security programmes across regulated environments, I have repeatedly seen organisations receive clean audit opinions while quietly carrying material exposure they had already identified but had yet to close. The audits were not wrong, and the auditors were not negligent. They were simply answering a different question from the one the board needed answered.

In one global rationalisation effort I led, we consolidated multiple legacy DLP and CASB controls into a modern, well-architected stack. The audit closed clean. Dashboards looked healthy. The board was satisfied. When we asked a simpler question, 'where is our most sensitive data actually moving, who is accessing it, and what are they doing with it', we found we did not have a good answer. The control surface was cleaner than it had ever been. The data exposure underneath it had quietly grown.

This is what I tend to describe as the Assurance Paradox. Boardroom confidence rises faster than the underlying protection, and the gap between what leadership believes is protected and what is exposed opens up quietly.

Audits validate that controls are defined, that they have been implemented, and that documentation exists to evidence both. That work matters, it creates discipline within an organisation, and it gives regulators a defensible mechanism for accountability.

What audits do not validate is whether those controls protect the most material data paths under real operating conditions, whether their underlying assumptions still hold once the business is under pressure, or whether the organisation would detect a breach in time to limit the blast radius.

None of this is a criticism of the audit profession. It is simply the design boundary of a sample-based, evidence-led exercise against a pre-defined control set, rather than a risk-led approach grounded in continuous validation, red-team testing and threat-informed frameworks such as MITRE ATT&CK.

2. The governing operating model has shifted underneath us

The environment that most audit and assurance frameworks were originally designed for, relatively static estates where server and database locations are known and enterprise usage is predictable, no longer exists in any large enterprise I work with. This has been driven in part by cloud-first initiatives, accelerated more recently by AI adoption.

This shift is now formally recognised at policy level. The UK's NCSC, in its 'Impact of AI on cyber threat: now to 2027' assessment, judges that AI is highly likely to increase the volume and impact of cyber intrusions through 2027, and the April 2026 open letter from the Secretary of State to UK business leaders made the same point in plain terms. AI-driven cyber risk is now an explicit board-level concern.

Sensitive data now sits across SaaS platforms, multi-cloud services, collaboration tools, partner ecosystems and a rapidly expanding set of AI services, each with its own identity model, telemetry surface and operational boundary. Identity has quietly replaced the network as the operational control plane, and rightly so given zero trust and identity being the new trust boundary. However, that shift has profound implications to how we can control sensitive data. In most enterprises today,

non-human identities (NHIs), such as service accounts, API credentials, application principals, and now AI agents, already outnumber human identities by something close to one hundred to one, up sharply from fifty to one only a year ago, and on a trajectory that shows no sign of slowing. Environments drift continuously. Temporary integrations become permanent fixtures, shadow applications arrive without anyone formally approving them, and configurations shift week by week without anyone reviewing the cumulative effect.

Underneath this technical reality sits an organisational one that is, in many ways, just as significant. Typically, the CIO, CDO and CTO are measured more on value generation and delivery velocity, while the CRO, CISO and DPO on risk reduction and regulatory compliance, with the boundary between the two groups often blurred. These two groups are pulling in different directions, and the data itself sits awkwardly in the middle, being reused, transformed and exposed faster than any static assurance view can reasonably track. Without knowing who is accountable. By the time an annual assurance report is signed off, the estate it describes has often moved on.

3. Why AI, and agents in particular, change the equation

AI did not start the data security problem, but it has fundamentally redefined the landscape. It has exponentially increased the risk of exposure, broadened its blast radius, and surfaced operating decisions that legacy controls were never built to mediate.

Modern AI systems generate far more data reuse, boundary crossing and automated decision-making than the systems they replaced. The intermediate artefacts they produce, including prompts, embeddings, retrieved context, outputs and audit logs, all carry sensitivity that most existing policy frameworks have not properly accounted for. Ownership of these artefacts becomes blurred almost immediately. Who owns the prompt entered into an enterprise GenAI assistant, the output that has been shaped by retrieval across three different data stores, or the embeddings that now sit in a vector database alongside thousands of others?

The sharper shift, and the one that most concerns experienced CISOs now, is agentic AI. Autonomous agents now take actions on behalf of users and systems, with their own identities, their own permissions, and in many enterprises, their own standing access to critical systems and data.

> **The four assumptions agentic AI breaks**
>
> 1. Stable roles
> 2. Predictable behaviour
> 3. Long-lived credentials
> 4. Clear human accountability
>
> *Agents break all four, often within their first few weeks of deployment.*

The pattern of recent OAuth-based compromises involving third-party AI tools illustrates this directly. Broad standing access granted once, rarely reviewed, becomes the path of least resistance for attackers and an unrecognised channel for data exposure.

The data backs this up in stark terms. The IBM Cost of a Data Breach Report 2025 found that of organisations which suffered an AI-related breach, 97% had no proper AI access controls in place, and 63% of breached organisations either had no AI governance policy in place or were still developing one. Most organisations are only beginning to grapple with what that means for their identity governance models, let alone the overall security posture including data governance.

The structured data blind spot offers a useful view of how slowly governance tends to move relative to operational reality. Many organisations still assume that structured data is adequately protected by encryption at rest, and the policy frameworks that govern it are largely built on that assumption.

Encryption at rest protects against storage compromise, which is a real and important threat. It does little, however, to address what happens when that data is queried, joined across databases, exported into a report, replicated into a SaaS analytics tool, or fed into an AI pipeline that generates derivative artefacts.

In hybrid estates, structured data moves substantially faster than the policies meant to govern it, and the lag between the two is where most of the actual exposure quietly accumulates.

The same logic applies to unstructured data, only at greater scale. Generative AI processes content in fragments, often referred to as chunks. Meaning and sensitivity can both be lost when those fragments are detached from their original context, and mixing content from multiple sources within a single context window can change the risk profile of an interaction in ways that no traditional DLP engine was designed to evaluate.

4. Where the failures actually happen

Security programmes are typically structured around key control points across identity, endpoint, network, application and data. Strong controls in each of these layers remain essential. Data exposure incidents, however, do not always happen at the control points themselves. They also happen in the paths between them, particularly where data is in motion or being reused outside its original context, and the data layer is where the consequences of any earlier control failure ultimately land.

Sensitive data is created in one context, accessed in another, reused for a different purpose, and then shared repeatedly, well past what was originally intended. Each transition introduces a fresh set of assumptions about intent, monitoring, ownership and containment, and those assumptions are rarely revisited once the initial design has been signed off. The principle of 'shift left', catching

issues at design and development rather than at runtime, applies to data paths just as it does to code. The earlier the assumption is tested, the cheaper and more contained the fix will tend to be. AI magnifies this dynamic by compressing the timescales involved. The controls themselves may still be present and well-configured, but the operating reality is now moving at a pace the original model was not designed to handle.

When an incident does occur, that path-level knowledge determines how quickly the business can isolate the blast radius, contain the financial and regulatory exposure, and resume operations on what is genuinely safe. Without it, recovery becomes a longer, more expensive exercise in working out what was actually affected.

What makes the difference, in practice, is context. Without it, automation simply becomes faster at magnifying false positives and errors at scale. To operate safely at any meaningful scale, every access and usage decision needs four pieces of context to be reliably available:

- what the data is, and how sensitive it is in the specific situation at hand;
- where it came from, and how it has been transformed or combined since it was first created;
- who or what is requesting access, and for what genuine purpose;
- and what action is appropriate given all of the above, at the point the decision is made.

If that context is not consistently available, or if it exists only in pockets within specific tools at specific control points, the organisation is, in effect, relying on policy by assumption. This is where false assurance is created.

5. False assurance

False assurance is what happens when an organisation can demonstrate compliance on paper but cannot demonstrate protection in practice.

The next twelve to eighteen months will compound this. DORA is now live, the EU AI Act is moving through phased enforcement, and regulated health industries are absorbing AI-specific compliance that, on paper, looks like reassurance. In my view, the opposite is more likely. Each new framework will produce its own audit cycle, its own attestations, its own dashboard. Each one is also a high-risk moment for false assurance, because most organisations will mistake regulatory readiness for operational protection. The frameworks are a floor, not a ceiling. Treating them as the answer is precisely the comfortable mistake this paper is about.

False assurance tends to appear when controls are present, audits pass and dashboards look healthy. On closer scrutiny, leadership cannot explain where exposure actually sits within the business, or which controls are working end to end on the most sensitive data paths.

The agreeable nature of false assurance is part of why it persists. It inadvertently lets executives defer the difficult cyber risk conversations that would have driven the right risk-based investment. The IBM Cost of a Data Breach Report 2025 places the current global average at $4.44 million per incident, rising to $10.22 million in the US, with a further $670,000 premium where shadow AI was involved. That is what false assurance costs in practice.

6. The board questions worth asking

The questions that shift a board-level security discussion from posture to assurance are not, in themselves, complex. They are uncomfortable largely because most organisations cannot answer them cleanly:

- Which data paths are most material to the business, not simply the most heavily regulated?
- Where are we relying on assumptions rather than evidence? 'DLP covers it' is comfortable to say. How do we actually know it does, in operational terms?
- Do we have a maintained inventory of sensitive data that tracks where it is copied, transformed and shared, rather than where it was originally classified?
- Where does AI reuse data outside its original intent, and how are we governing the prompts, retrieved context and outputs that result?
- Which non-human and agentic identities currently have access to sensitive data, and how does their actual usage compare with their assigned permissions?
- Which end-to-end scenarios have we actually tested, and which have we never tested?
- How do we detect drift, in access, in policy, in data movement and in AI behaviour, before that drift becomes an incident?

A clean answer to any three of these tends to put an organisation in the top quartile of its peer set. A clean answer to all seven is rare enough especially given the increasing adoption of AI.

7. What good looks like

Good assurance is not perfection, and it is rarely comfortable. It is a defensible answer to the question of whether an organisation can protect what truly matters, in the operating model it actually runs rather than the one its policies describe.

Six shifts make the difference in practice:

1. Make data paths explicit

Documentation that captures how data actually moves, reviewed after every meaningful change, not held as tribal knowledge by a select few.

2. Operationalise stewardship

Data ownership that survives reorganisations, leavers and matrix changes. Build a steward community that ensures ownership and unblocks any conflicts.

3. Treat AI artefacts as data

Prompts, intermediate artefacts, outputs and AI logs handled with the same lifecycle and retention discipline as any other sensitive data.

4. Govern NHIs and agents explicitly

Machine and agent identities with their own lifecycle, access model and audit trail, not treated as just a subset of human identity.

5. Move to continuous validation

Annual posture snapshots replaced by scenario-based testing of the highest-impact paths, end to end, against the operating model.

6. Build context into decisions

Context-aware controls including purpose, sensitivity, identity, session, destination. So scale becomes a multiplier of value, not of risk.

Taken together, these six shifts are not a list of practices. They are a single assurance system, each shift dependent on the others to be genuinely effective. This model takes its lead from NIST AI RMF, CSA AI Security frameworks, and the emerging Agent Access Management discipline. It translates their direction into the operating shifts most organisations now need to make.

The shift underneath all of these moves is the same:

Policy-only intent	Repository-level controls	Data path visibility	Continuous enforcement
Defined controls, but enforcement and visibility largely assumed.	Controls applied to known data stores; gaps between them remain.	Sensitive data tracked across systems, identities and AI artefacts.	Policy, telemetry, runtime and feedback operating as one layer.

Maturity over time →

From policy-only intent, through repository-level controls, to genuine data path visibility, and ultimately to proactive continuous enforcement, where policy, telemetry, runtime decisions and feedback loops work together rather than as separate layers managed by separate teams.

Lifecycle-led security models are emerging as one of the more credible operational answers to this challenge. I will explore them further in a future paper.

Final takeaway

With the adoption of AI, confidence in compliance without the right evidence is no longer a defensible position.

Organisations that avoid a headline cyber security breach will not be those with the most detailed control list, the most certifications, or the widest security tooling estate. They will be the ones that truly understand how sensitive data moves through their business, and can demonstrate with evidence which controls protect it where it actually matters. They will know this irrespective of whether they have been breached, and be ready to recover promptly if they are.

The shift is straightforward to understand and far harder to execute: from 'are the controls present?' to 'are we assured, end to end, on the paths that matter?'. A small but growing number of organisations are further ahead on this shift than their leadership realises. Most are further behind than their audit reports suggest. The choice, in the end, is no longer between more investment or less. It is between assurance that is real and assurance that is comfortable, and the next twelve months will, in my view, separate the two more decisively than the last twelve did.

About

I am a Cybersecurity practitioner and advisor with over twenty years of experience across consulting, system integrators, security vendors and in-house global & regional roles, leading data security initiatives and strategy. My work has spanned highly regulated industries including financial services, healthcare and pharma, and energy, with a global remit across hybrid cloud environments. I write and advise on the questions raised in this paper through DAISEQ Advisory, alongside senior leadership engagements in data, cloud and AI security.

This paper is intentionally my practitioner's view rather than a research report. If any of it resonates, or if you disagree, I would welcome the conversation.

Brian Wawengkang
United Kingdom
brianrw@daiseq.com

References

- Thales. 2026 Data Threat Report.
- Cloud Security Alliance and Oasis Security. Non-Human Identities and AI Security (2026).
- IBM Security and Ponemon Institute. Cost of a Data Breach Report 2025.
- National Cyber Security Centre (NCSC). Impact of AI on cyber threat: now to 2027 (2025).
- NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0).