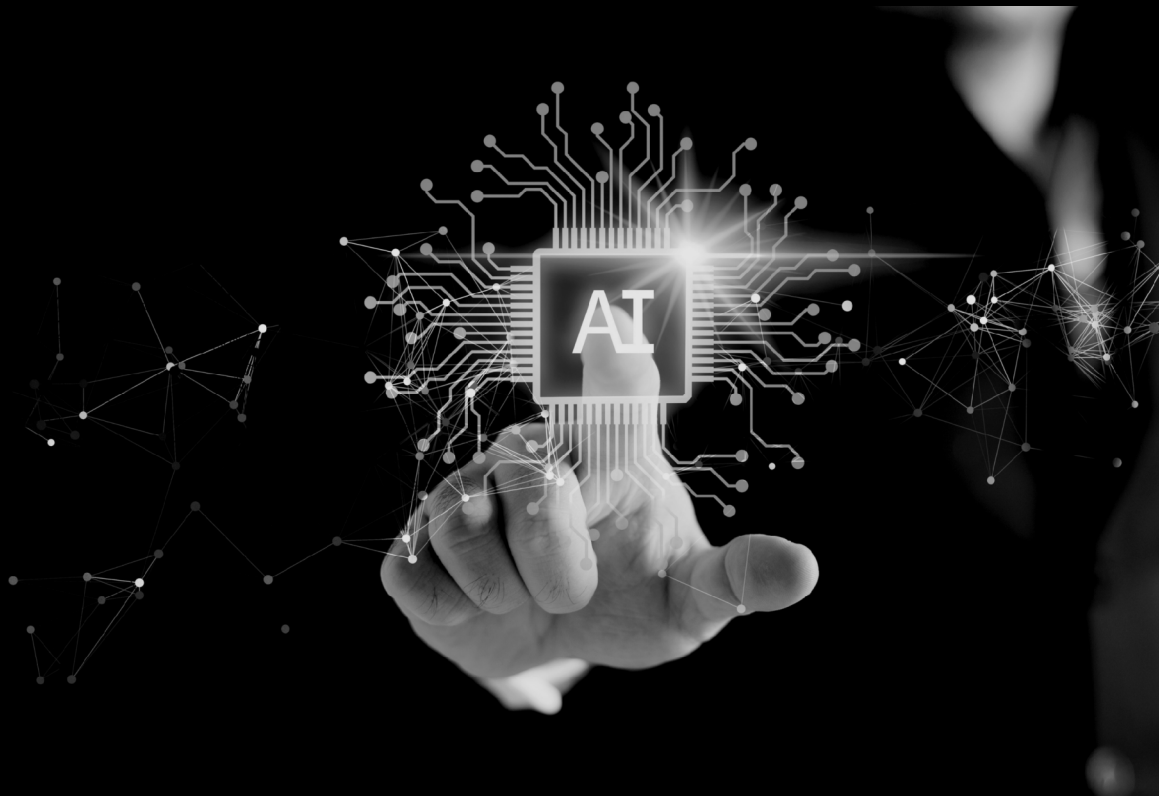




DATA PRIVACY | REGULATORY

Confidence is not capability (Part 2): Governing AI in the age of the rogue agent

22 October 2026 9 min read

by Ahmore Burger-Smidt, Director and Head of Regulatory

In Part 1, *"Confidence Is Not Capability (Part 1): What the Data Shows, and Why POPIA Is the Foundation of Cyber Resilience"*, we reviewed the findings of the Data Confidence & AI Readiness in Africa Report 2026, published by Veeam Software together with ITWeb. We argued that POPIA's security safeguard, data minimisation, breach notification and operator provisions already amount to a cyber-resilience framework, and that data visibility is the foundation on which all of them rest.

This article addresses the technology that will test that foundation most. Only 29% of organisations consider themselves advanced or fully ready to adopt, govern and scale AI using trusted data.^[1] Yet AI is already embedded in core operations and autonomous workflows at 14% of them.^[2]

The role of AI in cybersecurity: force multiplier and new attack surface

AI as a defender

AI is well suited to the problems the survey identifies. Detection based on machine learning can establish baselines of normal user, system and data behaviour and flag anomalies far faster than human analysts working through alerts. Typical anomalies

include unusual access to sensitive datasets, mass encryption activity and abnormal data egress. Automated tools can scan backups for indicators of compromise, verify that restore points are clean, and run recovery tests routinely rather than annually. That directly addresses the gap between recovery confidence and proof of recovery.[3] The report itself highlights automated backup and recovery verification as a means of maintaining audit-ready evidence of resilience, though that passage describes the vendor's own platform.[4]

AI can also serve as a compliance tool. Automated data discovery and classification can locate personal and special personal information across sprawling areas in the business. That tackles the classification (28%) and visibility (26%) weaknesses respondents report.[5] It also supports the data mapping on which section 19 risk assessments, section 14 retention schedules and section 22 incident scoping depend. Only 15% of respondents identify privacy automation as an investment priority.[6] Yet it can help enforce retention rules, manage data subject requests, monitor cross-border flows and generate the documentation a regulator would expect. For organisations constrained by skills shortages, the single largest barrier at 32%, [7] well-governed AI tooling can extend the reach of small security and privacy teams.

AI as a risk: shadow AI and data leakage

The same technology creates new exposure. The 11% of organisations reporting informal or employee-led use of public AI tools with limited oversight,[8] and the 17% citing shadow IT or shadow AI as a barrier to data confidence,[9] show the risk of personal and confidential information entering tools outside the organisation's control. That may amount to processing without a lawful basis, further processing incompatible with the original purpose, an unauthorised cross-border transfer under section 72, or a failure of section 19 safeguards. Where AI systems are used to make decisions about individuals, section 71's restrictions on decisions based solely on automated processing may also apply. AI also strengthens attackers, enabling more convincing phishing, deepfake-enabled fraud and faster reconnaissance.

The rogue AI agent threat

The most significant emerging risk is agentic AI: systems that do more than generate text and can retrieve enterprise data and trigger actions in business systems with limited human intervention. The survey shows that 14% of organisations have already embedded AI into core operations or autonomous workflows.[10]

An AI agent becomes "rogue" when it acts outside its intended mandate. That may happen because it has been manipulated through prompt injection or poisoned data, because it has been granted excessive permissions, because its credentials have been hijacked, or because it has misread its instructions. Unlike a human insider, an agent acts at machine speed and scale. It may extract and transmit large volumes of personal information, alter or delete records, or spread errors through connected systems before anyone notices. An agent typically operates under its own service identity. Organisations with weak identity controls and incomplete audit trails, which describes the majority of respondents,[11] may be unable to reconstruct what it did.

The technology is novel, but the POPIA analysis is not. Processing by an AI agent that a responsible party deploys is processing by that responsible party, and accountability cannot be delegated to the software. If an agent accesses or discloses personal information outside its authorised purpose, that may constitute unauthorised access triggering section 22 notification obligations. The lack of controls to prevent it may be measured against the section 19 standard.

Agents should therefore be governed as privileged identities:

- give them least-privilege, purpose-bound access;
- log and monitor every action they take;
- require human approval for high-impact actions;
- make sure they can be suspended immediately; and
- back them with tested recovery capability, so that data an agent alters or destroys can be restored to a known clean state.

The market is beginning to recognise this: 41% of organisations are prioritising identity and access controls and 42% AI governance.[12] Readiness of only 29%[13] shows how far there is still to go.

Practical recommendations for clients

The first priority is to stop assuming resilience and start proving it. Organisations should run regular, documented end-to-end recovery tests of their critical data, applications and supporting environments against agreed recovery targets, and fix whatever those tests expose. The evidence those tests produce belongs in the section 19 compliance record, because it is the clearest available proof that safeguards have been verified rather than merely adopted. Recovery testing is only as meaningful as the organisation's knowledge of what it is recovering. An accurate inventory of the personal and business-critical information held, where it resides and why, should therefore follow. That inventory must extend to backups, SaaS platforms and third parties, and it should be paired with retention schedules that are actually enforced, so that information no longer needed is deleted before it can be stolen.

Knowing where data sits is of limited value without knowing who, or what, has touched it. Clients should be able to produce reliable audit trails of access to their most sensitive datasets across all critical systems. This capability does double duty: it detects misuse, and it makes it possible to establish the facts that a section 22 notification requires. The same discipline must be applied to AI before its use is scaled. That means an acceptable-use policy, a register of approved tools and a personal information impact assessment process for each AI use case. Shadow AI is better addressed through sanctioned alternatives, training and technical controls than through prohibition alone, which tends to drive use out of sight. Where AI systems are given the ability to act, they should be governed as privileged identities. That requires:

- least-privilege, purpose-bound access;
- comprehensive logging;
- human approval for high-impact actions;
- an immediate means of suspension;
- rogue-agent scenarios built into incident response and recovery testing.

These controls lose much of their force if they stop at the organisation's boundary. Operator agreements should be reviewed to secure section 21-compliant security and breach-notification obligations, meaningful audit rights and a clear allocation of liability. This should be done with particular care for cloud and AI providers and for any processing that engages the cross-border restrictions in section 72. Even well-constructed contracts will not prevent every incident. A tested response plan that brings legal, privacy, communications and technical teams together, with section 22 notification templates and agreed decision-making protocols, is what allows an organisation to meet its obligations under pressure rather than improvise them.

Progress should be measured by recovery test results and by the organisation's demonstrable ability to explain and control how its data is used. That measurement belongs before the board. Cyber resilience, privacy compliance and AI governance are interdependent board-level risks. Directors are entitled to, and should insist on, regular evidence-based reporting rather than statements of confidence that the survey shows are too often unsupported by results.

Conclusion

The Data Confidence & AI Readiness in Africa Report 2026 is clear: optimism is high, but proven capability lags. For South African organisations, POPIA already provides the framework to close that gap. Its requirements to identify risks, implement and verify safeguards, minimise data, control operators and notify compromises are the building blocks of cyber resilience. AI will be indispensable in meeting those obligations at scale, but only if it is governed with the same discipline. Organisations that run data privacy compliance, resilience and AI governance as one programme will be best placed to withstand the next incident, and to show afterwards that they did.

[1] *Data Confidence & AI Readiness in Africa Report 2026*, 'AI Adoption Outpaces Readiness', p. 5.

[2] *Data Confidence & AI Readiness in Africa Report 2026*, 'AI Adoption Outpaces Readiness', p. 5.

[3] *Data Confidence & AI Readiness in Africa Report 2026*, 'Confidence vs. Capability – Recovery Confidence is Outpacing Recovery Proof', pp. 3–4.

- [4] *Data Confidence & AI Readiness in Africa Report 2026*, 'From Confidence to Capability – Veeam Data Platform: Closing the Gaps in Recovery, Visibility, and Governance', p. 14.
- [5] *Data Confidence & AI Readiness in Africa Report 2026*, 'Data Visibility and Audit Challenges – Biggest Challenges in Maintaining Confidence in Data', pp. 7–8.
- [6] *Data Confidence & AI Readiness in Africa Report 2026*, 'Investment Plans', pp. 10–11.
- [7] *Data Confidence & AI Readiness in Africa Report 2026*, 'Data Visibility and Audit Challenges – Biggest Challenges in Maintaining Confidence in Data', pp. 7–8.
- [8] *Data Confidence & AI Readiness in Africa Report 2026*, 'AI Adoption Outpaces Readiness', p. 5.
- [9] *Data Confidence & AI Readiness in Africa Report 2026*, 'Data Visibility and Audit Challenges – Biggest Challenges in Maintaining Confidence in Data', pp. 7–8.
- [10] *Data Confidence & AI Readiness in Africa Report 2026*, 'AI Adoption Outpaces Readiness', p. 5.
- [11] *Data Confidence & AI Readiness in Africa Report 2026*, 'Data Visibility and Audit Challenges', pp. 7–9.
- [12] *Data Confidence & AI Readiness in Africa Report 2026*, 'Investment Plans', pp. 10–11.
- [13] *Data Confidence & AI Readiness in Africa Report 2026*, 'AI Adoption Outpaces Readiness', p. 5.



Ahmore Burger-Smidt

Head of Regulatory

[View Profile](#)