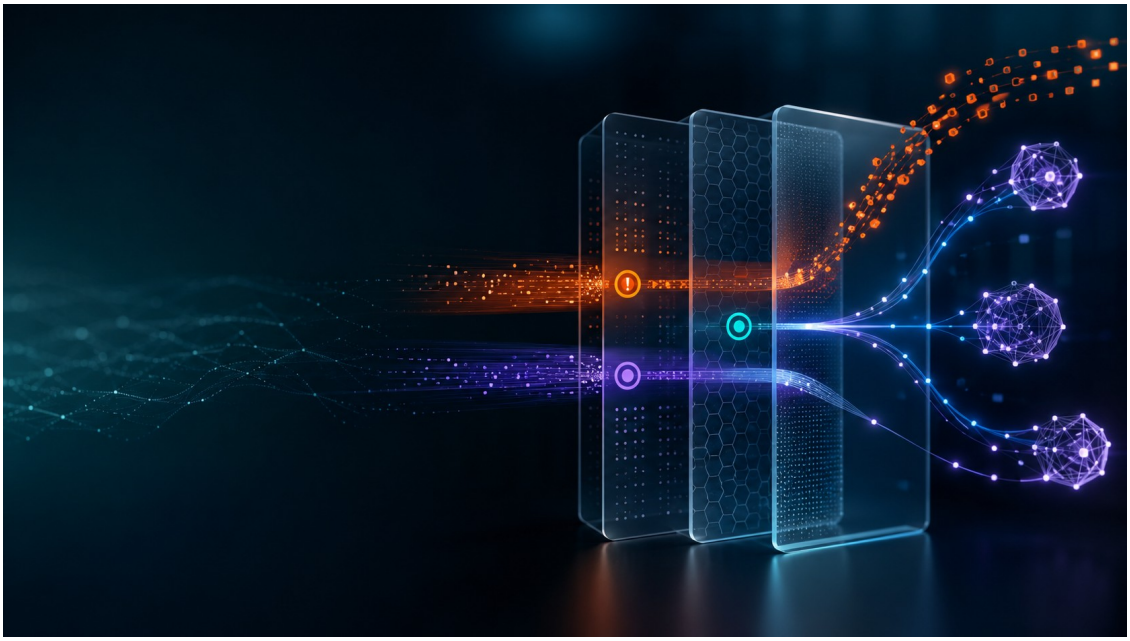
SOTERIA CLOUD WHITE PAPER

Governing generative AI without stopping the business

A practical operating model for Shadow AI visibility, prompt data protection, harmful-prompt controls and accountable managed services.



Generative AI is becoming an everyday business tool. Governance must therefore move beyond a one-time acceptable-use policy and become a living operating model that observes use, makes risk-based decisions, enforces controls and produces evidence.

Audience: executives, Information Officers, security and IT leaders, MSP owners, compliance teams and service managers.

Executive summary

The central GenAI risk is not the technology alone. It is uncontrolled interaction between users, sensitive information, external AI services and weakly defined business policy. Acronis GenAI Protection addresses this through visibility, prompt-level data loss prevention, harmful-prompt protection, application access control and reporting within a multitenant MSP platform.

A balanced objective

The goal is not to stop employees from using AI. It is to create a safe path for approved use, identify unapproved or risky activity and give the organisation enough evidence to improve its decisions. That requires policy, technical controls, business ownership and user enablement to work together.

The operating model in one line

- Observe actual use before enforcing assumptions.
- Decide which tools, purposes and data categories are acceptable.
- Enforce proportionate controls at the point of interaction.
- Review evidence, exceptions and outcomes on a defined cadence.

> **Core conclusion:** A secure AI programme should enable productivity and constrain risk at the same time. The measure of success is not simply the number of blocked prompts; it is whether approved AI use becomes more visible, safer and easier to govern.

A four-part GenAI risk model

Organisations often focus only on confidential data leakage. In practice, GenAI adoption introduces four connected governance problems.

Risk domain	Typical scenario	Governance response
Shadow AI	Teams use unapproved browser tools outside IT visibility	Discover services, attribute use and create an approved application list
Sensitive-data leakage	A user pastes personal information, credentials or confidential content into a public AI tool	Classify prompt data and block unauthorised submission
Prompt abuse	A prompt attempts to manipulate AI behaviour, bypass rules or introduce unsafe content	Detect harmful patterns and enforce the approved-use policy
Governance fragmentation	Policies, exceptions, incidents and reporting sit in different systems with unclear ownership	Centralise evidence, roles, review and service reporting

The control architecture: observe, decide, enforce, evidence

Acronis GenAI Protection can be understood as four operational layers rather than a single blocking feature.

Risk differs by use case

A marketing team drafting generic copy, a developer reviewing proprietary code and a finance user analysing customer records do not present the same risk. Policy should consider the user, business purpose, information class, AI application, contractual terms and required human review.

Layer	Acronis capability	Management decision
Observe	Shadow AI monitoring, user attribution, events, widgets and reports	What is being used, by whom, and where is the concentration of risk?
Decide	Application-level allow or block policies and configured sensitive-data categories	Which applications, domains, data and purposes are approved?
Enforce	Prompt DLP, harmful-prompt detection and block mode	Which violations must be stopped automatically and which should generate evidence only?
Evidence	Central event log, trends and multitenant reporting	How will exceptions, incidents, executive reporting and continuous improvement work?

> **Design principle:** Start with the business decision and then configure the control. A technically possible block is not automatically a good policy; it needs a defined owner, rationale, exception path and review date.

Scope is important

Acronis describes GenAI Protection around browser-based GenAI interactions. Organisations needing protection across additional local and network channels, including local GenAI applications, should assess the broader Acronis DLP service and state the coverage difference clearly.

Protect the prompt and govern the application

The highest-value controls act before sensitive or harmful content leaves the user's workflow.

Prompt data loss prevention

GenAI Protection can inspect prompts for regulated and business-sensitive categories such as personally identifiable information, health information, payment-card-related data, credentials and content marked confidential. In block mode, unauthorised submissions can be stopped before they reach public or unsanctioned AI services.

Harmful-prompt protection

Prompt injection and other abusive techniques can attempt to manipulate model behaviour, bypass intended rules or introduce unsafe downstream content. Detection and blocking reduce exposure, but output validation and human accountability remain necessary for material decisions.

Application access control and user attribution

The 26.07 release added application-level policy modes: allow only specified applications and domains, or block specified applications and domains while permitting the rest. The same release added logged-in user attribution to GenAI events, closing the gap between device visibility and accountable user activity.

Move from detect-only to enforceable governance

A staged rollout lowers disruption and improves policy quality.

1. **Baseline.** Run detect-only mode long enough to identify applications, departments, users, data patterns, false positives and legitimate business demand.
2. **Classify.** Group AI services as approved, conditionally approved, unapproved or under review; map data categories and business purposes.
3. **Assign ownership.** Name the business policy owner, Information Officer or privacy lead, technical operator, exception approver and incident escalation path.
4. **Enforce proportionately.** Block high-confidence sensitive-data and harmful-prompt events first; apply application controls where the approved list is mature.
5. **Review and improve.** Use events, widgets and reports to tune policy, coach teams, close coverage gaps and report business outcomes.

Evidence should drive the service review

- Top AI applications, users and departments by activity.
- Sensitive-data attempts by category, trend and disposition.
- Harmful-prompt events and policy violations.
- Blocked, allowed, overridden and unresolved events.
- Exceptions due for review and approved applications with changed terms or risk.

POPIA-aware AI governance and managed service design

POPIA does not name a product that makes AI compliant. It establishes conditions for lawful processing and accountability. Technical controls can support those objectives when they are embedded in a broader governance programme.

Governance objective	Practical AI-control response	Caution
Accountability and purpose	Define approved AI purposes, owners, user groups and applications	A technical allow-list does not prove lawful purpose
Processing limitation	Reduce unnecessary prompt data and block prohibited categories	Classification must be tuned and reviewed
Security safeguards	Use prompt DLP, harmful-prompt controls, monitoring and incident workflows	Coverage and service boundaries must be documented
Operator governance	Clarify the roles of the customer, MSP, Soteria Cloud, Acronis and AI service providers	Contracts, cross-border processing and breach duties require legal review
Openness and user practice	Publish an AI-use standard, explain monitoring and train users	Covert or unclear monitoring can undermine trust and governance

Suggested managed-service tiers

- Discover: detect-only baseline, application inventory and initial risk report.
- Govern: approved-tool policy, application access control, user guidance and exceptions.
- Protect: prompt DLP, harmful-prompt enforcement, incident triage and policy tuning.
- Assure: executive reporting, quarterly governance reviews, broader DLP alignment and continuous improvement.

> *This white paper is general information, not legal advice. Organisations should obtain appropriate legal, privacy, labour and contractual guidance for their circumstances.*

A 90-day implementation roadmap

The first objective is not maximum blocking. It is a stable control model that the business understands and can operate.

1. **Days 1-30: discover.** Confirm stakeholders and scope, deploy monitoring, establish the baseline, identify top tools and sensitive-data patterns, and draft the approved-use standard.
2. **Days 31-60: govern.** Approve application categories, tune sensitive-data rules, define exceptions and escalation, pilot user communications, and test reporting.
3. **Days 61-90: enforce.** Activate agreed blocks, review event quality, close operational gaps, publish the service report and schedule the next policy review.

Measures that matter

- Percentage of observed AI usage occurring in approved tools.
- Sensitive-data attempts prevented, investigated and repeated.
- False-positive and exception rates by rule.
- Time from event to review and policy decision.
- User awareness, adoption of approved tools and business-owner satisfaction.

Start with evidence, then enforce

Soteria Cloud can help structure the discovery period, policy model, control rollout and partner-ready managed service.

Ask AB about GenAI Protection

Sources and product notes

1. Acronis GenAI Protection product page - official source
2. Acronis GenAI Protection data sheet - official source
3. Acronis GenAI Protection general availability announcement - official source
4. Acronis Cyber Protect Cloud 26.07 release notes - official source
5. Acronis GenAI Protection data loss prevention walkthrough - official source
6. Acronis GenAI Protection dashboard and reporting walkthrough - official source
7. Protection of Personal Information Act 4 of 2013 - official source
8. South African Information Regulator - POPIA - official source
9. South African National AI Policy Framework - official source

Sources reviewed 24 August 2026. Acronis product capability, licensing and availability may change. Verify current official documentation before publication or contracting.