AI SECURITY WHITEPAPER

Beyond posture: AI security for the agentic era

Unified AI security and governance

relyanceai

Table of Contents

1. Why AI Creates a Never-Before-Seen Risk Landscape

AI isn't just "software with data" – it's software whose behavior is continuously shaped by data. Unlike traditional applications, where logic is fixed in code, AI systems generate decisions, actions, and content in real time based on incoming inputs. This creates a risk surface unlike anything CISOs or privacy teams have had to secure before.

Data as executable logic

In AI systems, the data isn't just used – it's run. A customer record, a support ticket, or a workflow log can directly shape model outputs, prompt behavior, or trigger downstream system actions. When data becomes executable, the threat model changes fundamentally: unauthorized or manipulated data entering an AI system can alter what the system does, not just what it knows.

Behavioral fluidity

AI doesn't follow pre-coded control paths. The same model, on the same endpoint, may behave differently from one moment to the next depending on inputs and context. This makes traditional posture management – which assumes stable, inspectable configurations – insufficient. Behavioral drift is a security event, not just a performance concern.

Agentic autonomy

With the rise of agentic AI, models aren't limited to responding – they take action, chain tools, call APIs, spawn sub-agents, and iterate on plans. A compromised or misconfigured agent with administrative access to sensitive data isn't just a misconfiguration issue; it's an insider threat operating at machine speed.

Top AI security risks for enterprise AI adoption

01 Sensitive Data Leakage

Exposure of regulated data through AI training or inference flows.

02 Autonomous Agent Overreach

AI agents chaining tools in unsafe, unintended ways – escalating privileges and accessing data far beyond intended scope.

03 Overprivileged AI Access

Excessive model permissions or weak IAM creating AI assets with unchecked access to critical systems.

04 Shadow & Unsanctioned AI

Unapproved models, agents, and integrations deployed without security review – invisible to existing tooling.

05 AI Supply Chain Compromise

Third-party models, MCP servers, and open-source components introducing vulnerabilities that propagate into production.

06 Compliance Drift

Runtime behavior diverging from regulatory controls – EU AI Act, NIST AI RMF, ISO 42001, OWASP – without visibility into the gap.

AI's risk surface is both behavioral and informational. Securing and governing it means tracking not only where data goes, but how it changes what the system does.

2. Security and Governance Realities of Enterprise AI

These risks manifest in specific, recurring problems that security and compliance teams face every day as AI adoption accelerates. They map to three fundamental gaps: incomplete visibility, insufficient risk context, and the inability to act with confidence.

Incomplete Visibility	No complete picture of what AI can reach and touch

The challenge

Every AI system that can touch sensitive data is a potential exposure. But most organizations have no reliable, current inventory of the AI systems in their environment. Models are deployed by engineering teams, AI features are enabled in SaaS tools, agents are built and run across the business – often without IT or security visibility. Security teams are making risk decisions based on an incomplete picture.

- **Shadow AI and unsanctioned tools:** developers integrate models, business users enable AI features in SaaS platforms, and vendors quietly turn on AI capabilities within existing contracts. These systems frequently process sensitive data – customer records, financial information, PHI – with no record of what flows where.
- **AI agents and MCP servers** as untracked identities: AI agents act as first-class identities in enterprise environments, calling APIs and accessing databases with real permissions. MCP servers are rapidly becoming the standard channel through which agents interact with enterprise systems. Most organizations have no automated process to discover either.
- **Third-party vendor AI** as a blind spot: SaaS vendors with AI capabilities enabled often operate without visibility or runtime validation, creating risk exposure that traditional vendor management processes can't catch.
- **Code and pipeline gaps:** cloud-centric tools miss risks that originate in source code and CI/CD pipelines – data flowing from a repository into a training pipeline, or credentials embedded in agent scaffolding – before they ever reach a cloud resource.

Lack of Risk Context	The most dangerous risks are invisible in isolation

The challenge

Knowing that a sensitive data store exists, or that an AI agent has broad permissions, is not the same as understanding risk. The most dangerous vulnerabilities in AI environments are compound: they form when individually low-risk elements converge into a lethal combination that no individual scan would flag.

Compound risk, defined

An AI agent alone may not be dangerous. A Snowflake table with customer PII may be properly secured. But an AI agent with administrative access to that table – via a service identity no one is monitoring – is a critical vulnerability. Not "you have sensitive data in Snowflake." Instead: "This AI agent with over-privileged access can reach customer PII through an unvetted MCP server, using a service identity no one is monitoring." That is the difference between a posture alert and actionable risk context.

- **Overprivileged agents with access to sensitive data:** the combination of agent identity, access scope, and data sensitivity is the risk. Assessing each dimension separately makes this invisible.
- **Identity-to-data relationships for non-human identities:** AI agents, service accounts, and automated pipelines often outnumber human users and carry significant access privileges. Tracing permission chains from agent to API key to service account to sensitive data requires connecting identity posture to data posture.
- **AI supply chain risks:** vulnerabilities in third-party models, MCP servers, and open-source components may not be dangerous in isolation but become critical when mapped to the sensitive data those components can reach.

No Guided Path to Resolution

Fragmented alerts with no connected story create an investigation tax

The challenge

Even when security tools surface findings, acting on them is slow. Most AI security tools generate alerts that require security teams to manually stitch together context from multiple systems – a SIEM, a CNAPP, an identity tool, a data catalog – before they can understand severity, scope, or what to do. The result is an investigation tax that consumes cycles and slows response.

- **Fragmented tooling means no integrated view:** security teams typically operate across four or more tools to understand a single AI risk finding. Each tool sees one dimension – the AI asset, the data, the identity, the infrastructure – and no single tool assembles the full picture automatically.
- **Posture alerts are not actionable risk context:** knowing that an AI agent "has broad permissions" is not the same as knowing what sensitive data it can reach, what path it would take, and what code change would close the exposure. Real actionable risk context is not alert noise.
- **Compliance evidence assembled retroactively:** the EU AI Act, NIST AI RMF, and ISO 42001 require documented evidence of AI behavior and risk management. Most organizations assemble this manually, querying multiple systems and producing documentation that reflects a past state rather than current reality. Time from risk identified to risk resolved can reduce by an order of magnitude when evidence is always current.
- **No governance continuity:** AI systems drift. Models get retrained, agents gain new tools, permissions change. Without continuous monitoring connected to live data and identity context, governance is point-in-time rather than operational.

3. Current Tools Don't Close These Gaps

The AI security market has expanded rapidly. Cloud security platforms have added AI-SPM features. DSPM vendors have added AI asset discovery. AI lifecycle security tools have emerged. Each addresses a need – but none provides the complete picture for AI security and governance.

Visibility Gaps	Cloud-centric and data-at-rest tools miss most of the AI attack surface

Cloud-native security platforms that have added AI-SPM capabilities start from cloud infrastructure posture and layer AI asset discovery on top. They discover AI running in cloud infrastructure – but miss code-level and SaaS-embedded AI agents. They have no visibility into source code repositories or CI/CD pipelines where risk often originates. SaaS-embedded AI features and shadow AI adopted by business users are outside their detection scope entirely.

Traditional DSPM tools discover and classify sensitive data in cloud storage and databases, but stop there. They scan data at rest – they don't follow data as it moves, track how AI agents access it, or connect it back to the source code where risk originates. They also do not trace data flowing into AI training pipelines or inference flows.

Lack of Risk Context

Posture awareness without compound risk detection leaves the worst threats invisible

AI-SPM tools focus on posture awareness – what AI assets exist and how they are configured. That is only the first step. They do not map the relationships between AI agents and the sensitive data they access, which means compound risks that emerge only at the intersection of data, identity, and agent behavior are invisible.

Data security platforms focused on classification and cloud data stores have limited identity intelligence and do not treat AI agents as first-class identities alongside human and service accounts. They cannot trace how sensitive data flows into AI models or through MCP servers, and do not surface the lethal combinations that form when overprivileged agents meet sensitive data.

No Actionable Explainability

Isolated alerts with no connected story force manual investigation across fragmented tools

AI-SPM and DSPM tools typically deliver posture alerts – findings that identify a state without providing the full context chain needed to understand severity, scope, and resolution path. Security teams must manually investigate across multiple tools to reconstruct what happened, who is involved, what data is at risk, and what to do. No single tool assembles the story. This is the investigation tax.

AI lifecycle security tools address important pipeline risks but do not monitor what AI systems are doing with data at runtime, and do not provide the continuous governance evidence that regulators increasingly require. Compliance evidence remains manually assembled and perpetually out of date.

The core gap

AI-SPM alone focuses on posture awareness – what exists and how it's configured. It does not provide the data context or remediation needed for complete AI security. Real AI security requires correlating identity, access permissions, and data sensitivity in real time to surface dangerous combinations that no individual tool can see. Relyance AI treats AI-SPM as a capability within its broader data defense platform.

4. Relyance AI: AI Security Built for the Agentic Era

Relyance AI delivers a unified AI Security Posture Management and AI Governance platform built on a data-journey-first architecture. Every AI data flow – from source code and pipelines to inference and agent actions – is continuously mapped. The result is complete visibility into every AI asset and data exposure, risk context that surfaces compound threats no individual tool can see, and actionable explainability that drives fast resolution from a single pane of glass.

See Every Data Exposure

Everything that can reach, touch, or interact with your sensitive enterprise data – mapped continuously.

Relyance AI delivers continuous, automated discovery of AI applications, models, datasets, MLOps workflows, and third-party AI integrations – including unapproved shadow AI – across training and inference workflows. Every AI asset in the path – models, code, pipelines, APIs, agents, MCP servers, third-party integrations – is continuously discovered and mapped. No blind spots. No partial views. The complete picture.

Key capabilities

- **Comprehensive AI asset discovery –** automated, agentless discovery across code, cloud, and SaaS including shadow AI and unsanctioned tools.
- **Sensitive data classification and mapping –** AI-powered detection of PII, PHI, PCI, and enterprise-specific data across the full AI lifecycle.
- **AI Data Journeys™ –** runtime and training-time tracing of how sensitive data enters, flows through, and exits AI systems, from source code through inference.
- **MCP server discovery –** automated discovery of first- and third-party MCP servers, mapped to the data they reach and the identities using them.
- **Shadow AI detection –** surfaces AI systems operating outside approved governance processes, including SaaS-embedded AI features enabled without IT awareness.

Top risks addressed

- Shadow AI and unsanctioned tools creating blind spots
- Sensitive data leakage through AI training or inference
- AI supply chain risks from third-party integrations
- MCP servers as unmanaged attack surfaces
- Code-level and pipeline risks missed by infrastructure-focused tools

Understand Compounding Risk

Every finding arrives with the full context-chain already assembled – what's wrong, why it's wrong, who's involved, what's impacted.

Relyance AI's Data Exposure Graph connects data sensitivity, identity privilege, access behavior, and AI agent actions into a single risk view. This is what enables compound risk detection –

surfacing the dangerous combinations that form when individually low-risk elements converge, and that are invisible to tools that assess data and identity posture separately.

Powered by the Data Journeys™ and Data Exposure Graph – connecting data, identity, and AI to surface risks that only emerge at the intersection.

Key capabilities

- **Identity-to-data intelligence –** maps how human users, service accounts, AI agents, and automated workflows gain access to sensitive data, tracing permission chains across the environment.
- **Non-human identity coverage –** AI agents and MCP servers treated as first-class identities alongside human users and service accounts.
- **Compound risk detection –** correlates data sensitivity, identity permissions, and access behavior in real time to surface dangerous combinations no individual scan would flag.
- AI supply chain risk context – maps third-party model and MCP server risks in the context of the sensitive data they can reach.

Top risks addressed

- Overprivileged AI agents with unchecked access to sensitive data
- AI agent identity and privilege chains invisible to infrastructure-focused tools
- Compound risks across data, identity, and AI behavior
- AI supply chain vulnerabilities with sensitive data exposure context

Resolve Fast with Integrated Risk Response

Real actionable risk context, not alert noise. No investigation tax. No stitching across tools.

Relyance AI delivers contextual remediation – not just findings, but the full chain assembled: what is wrong, who is involved, what data is at risk, and where in the code to fix it. For compliance teams, it means evidence always current from a live posture, not assembled retroactively. Lyo™ gives your team a natural language interface to interrogate risk and act – not navigate dashboards.

Key capabilities

- **Contextual remediation —** every finding comes with the full risk chain: what is wrong, who is involved, what data is at risk, and what to fix.
- **Continuous AI governance —** runtime policy mapping, model lineage and drift tracking, and a historical audit framework that stays current as AI systems change.
- **Live compliance mapping —** automated, continuous mapping to EU AI Act, NIST AI RMF, and ISO 42001 from a live AI inventory, not a retroactively assembled report.
- **Third-party vendor AI risk —** continuous assessment replacing point-in-time questionnaires, with DPIA enrichment from live data flows and contract validation against runtime behavior.
- **Lyo™ —** natural language interface to interrogate risk and act, not navigate dashboards.

Top risks addressed

- Investigation tax from fragmented, context-poor alerts
- Compliance evidence assembled manually and retroactively
- Governance that is point-in-time rather than continuous
- Regulator and audit inquiries requiring weeks of manual assembly
- AI drift creating undetected compliance and posture gaps

AI Governance business value

- Policy validation time cut by up to 60%
- Regulator inquiry turnaround reduced from weeks to hours

Vendor AI risk business value

- AI vendor risk assessment time cut by up to 70%
- Always-current view of vendor AI risk posture

What security leaders say

“AI is creating and moving data faster than any team can track. Only AI-native tools like Relyance AI can keep up.” – Chris Bender, CISO, ClickUp

“Most tools show me a snapshot. Relyance gives me the full movie, around the clock.” – Karthik Chakkarapani, SVP & CIO, Zuora

“Relyance gives us a complete, contextual understanding of our data landscape. We can see what we have, how it's used, who owns it, and where the risks are.” – Jason James, CIO, Aptos

Conclusion

AI security is not a feature to add to existing security platforms. It requires a fundamentally different architecture — one that starts from the data journey, maps behavior at runtime, connects AI risk to data risk, and delivers continuous evidence of compliance.

Traditional tools were built for deterministic applications. AI systems are non-deterministic, continuously shaped by data, and increasingly autonomous. Securing them means tracking not only where data goes, but how it changes what the system does.

Relyance AI was built for exactly this environment. Complete visibility into every AI asset and data exposure. Contextualized risk that surfaces compound threats no individual tool can see. And actionable explainability — real remediation context, a connected story assembled automatically, continuous governance, and compliance evidence always current. Not AI security bolted onto an existing product — AI security and governance built from the ground up for the agentic AI era.

Agentless, AI-native architecture

Relyance AI does not require installing sensors or software agents on servers. It integrates via APIs and reads from logs, source code repositories, cloud configuration, and other existing data sources. Most organizations can connect the platform and start seeing insights within minutes, not weeks.

The platform's TrustiQ™ intelligence engine uses advanced machine learning and natural language processing to automatically classify data – including unstructured data and code – and to infer the context of data processing activities. It can distinguish personal health information from log data, or recognize that an email address is being used for marketing versus authentication, based on how it flows through code. As new patterns emerge – a new cloud service, a new MCP server protocol – the AI models accommodate them without manual rule updates.

Deployment options

	Full SaaS	**InHost™ (VPC)**	**DirectConnect**
What it is	Cloud-hosted, multi-tenant SaaS on GCP. Each customer logically isolated via dedicated IAM roles and resource boundaries.	Entire platform runs within your own VPC. Sensitive telemetry never leaves your environment.	Private link between your internal network and the Relyance platform. No public internet exposure.
Best for	Rapid deployment with minimal operational overhead.	Strict data sovereignty and security requirements.	Environments where no public internet exposure is acceptable.

Security by design

- Data transmitted via encrypted channels (TLS 1.3)
- Architecture aligned with Zero Trust principles
  Role-based access controls and audit logs
- SOC 2 compliant
- No rip-and-replace: augments existing stack by integrating with developer tools, CI/CD pipelines, cloud accounts, and ticketing systems

Regulatory frameworks supported

EU AI Act	Risk-based classification of AI systems (prohibited, high-risk, limited-risk, minimal-risk); mandatory documentation; post-market monitoring for high-risk AI; transparency obligations.
NIST AI RMF	Data provenance and usage tracking; risk identification and management across the AI lifecycle.
ISO/IEC 42001	End-to-end traceability; quality and lawful use documentation for AI management systems.
GDPR / CCPA / CPRA	Data minimization; lawful basis documentation; right to erasure; breach notification obligations.
HIPAA / HITECH	Minimum necessary principle; security rule; breach notification for protected health information.
SOC 2 / NIS2 / SOX	Security and availability controls; sector-specific regulatory obligations.