

Gartner.

Licensed for Distribution

This research note is restricted to the personal use of ().

Market Guide for AI Trust, Risk and Security Management

18 February 2025 - ID G00785960 - 58 min read

By Avivah Litan, Max Goss, [and 4 more](#)

The AI trust, risk and security management market comprises four layers of technical capabilities that enforce AI governance policies. The top two layers, new to AI, are consolidating into a distinct market segment. AI leaders should use this guide to understand TRiSM and discover sample vendors.

Overview

Key Findings

- Enterprises face multiple AI risks and are most concerned with data compromise, third-party risks, and inaccurate or unwanted outputs:
 - Malicious hacks against enterprise AI are still uncommon, while incidents of unconstrained harmful chatbots are well documented and internal oversharing data compromises are prevalent.
- Layered AI TRiSM measures apply to all types of AI, from embedded AI to homegrown, agentic AI systems, and complement traditional technology protection measures:
 - New AI governance and runtime enforcement functions are coalescing to form a distinct, unique segment of the TRiSM market.
- User demand for GenAI TRiSM solutions is steadily increasing, and providers of all sizes are competing for this new enterprise business:

- Some vendors specialize in supporting security and risk mitigation, others in supporting ethical or safe practices and regulatory compliance. No single vendor or solution addresses all segments and aspects of AI threats and risks at this time.
- AI trust, risk and security issues surface organizational silo issues, pushing teams to realign to solve problems that cross departmental boundaries.

Recommendations

AI leaders and individuals responsible for AI controls must work with stakeholders to establish AI governance policies and assign responsibilities for managing AI TRiSM technology that supports policy execution. Once organized, they must:

- Discover and inventory all AI used in the organization.
- Revisit and implement data classification, protection and access management across all enterprise information that can potentially be utilized by AI.
- Evaluate and implement layered AI TRiSM technology to continuously enforce policies across all AI use cases.
- Stay flexible and ensure independence from any single AI model provider as more advanced and cost-effective AI models and tooling become available.
- Monitor the vendor landscape frequently for expected consolidation and emerging TRiSM outsourced services that reduce the burden of system integration and maintenance.

Strategic Planning Assumptions

Through 2026, at least 80% of unauthorized AI transactions will be caused by internal violations of enterprise policies concerning information oversharing, unacceptable use or misguided AI behavior rather than malicious attacks.

By 2027, AI TRiSM as a service will emerge as a viable outsourced service option for enterprises that do not have the resources to implement their own AI TRiSM services.

By 2028, consolidated and dedicated information governance teams consisting of representatives from data and analytics, digital workplace, and security and compliance will exist in 25% of large organizations, up from less than 1% in 2023.

Market Definition

Gartner defines the AI trust, risk and security management (TRiSM) market as comprising four layers of technical capabilities that support enterprise policies for all AI use cases. These policies help assure AI governance, trustworthiness, fairness, safety, reliability, security, privacy and data protection. The top two layers — AI governance and AI runtime inspection and enforcement — are new to AI and are, in part, consolidating into a distinct market segment, while the bottom two layers represent traditional technology focused on AI.

AI brings new trust, risk and security management challenges that conventional controls do not address. Enterprises are most concerned with data compromise, third-party risks, and inaccurate or unwanted outputs, and need to ensure enterprise AI behavior and actions align with enterprise intent.

Enterprises must further retain independence from any single AI model or hosting provider to ensure scalability, flexibility, cost control and trust, as AI markets rapidly mature and change. Enterprises need to be able to pick and choose which AI tools and models best fit their use cases from a performance, cost and risk point of view, and cannot simply rely on vendors adding AI functionality to existing enterprise applications to meet all the organization's AI requirements.

Enterprise AI TRiSM functions should be portable across back-end hosting providers. Enterprises should supplement, where they can, their own TRiSM services with TRiSM functions provided by hosting and frontier model providers to support a robust, layered control framework.

The AI TRiSM market has been steadily growing and evolving with solutions that address these issues. User demand for TRiSM solutions is increasing as more AI projects are operationalized, and providers of all sizes are competing for this new enterprise business.

AI TRiSM solutions give enterprises visibility into all of the AI that is accessed and used in their networks. They also enable organizations to more safely use AI, ensure AI actions align with organizational intent, keep AI systems secure from malicious actors, and assure confidential data and intellectual property are properly protected. AI TRiSM solutions also help support compliance with AI regulations, safety and security standards, and internal controls.

AI TRiSM measures apply to all types of AI, from embedded AI to homegrown, agentic AI systems, and complement traditional technology protection measures. New AI governance

and runtime enforcement functions are coalescing to form a distinct, unique segment of the TRiSM market.

Mandatory Features

Mandatory features include:

- AI catalog: An inventory of AI entities used in the organization, such as models, agents and applications in various configurations, ranging from embedded AI in off-the-shelf, third-party applications to bring-your-own AI, to retrieval-augmented generation systems, to first-party models
- AI data mapping: Data used to train or fine-tune AI models, provide context for user queries in a RAG system, or feed AI agentic systems
- Continuous assurances and evaluation of performance, reliability, security, or safety expectations and metrics that are used for baselines
 - Assurances and evaluations are applied predeployment, postdeployment (out of band)
- Runtime inspection and enforcement: Applied to models, applications and agent interactions to support transactional alignment with organizational governance policies. Applicable connections, processes, communications, inputs and outputs are inspected for violations of policies and expected behavior. Anomalies are highlighted and either blocked, autoremediated or forwarded to humans or incident response systems for investigation, triage, response and applicable remediation.

Market Description

AI TRiSM technology is quickly evolving to keep pace with rapid AI advancements and users' increasing interest in deploying it safely and securely. ¹

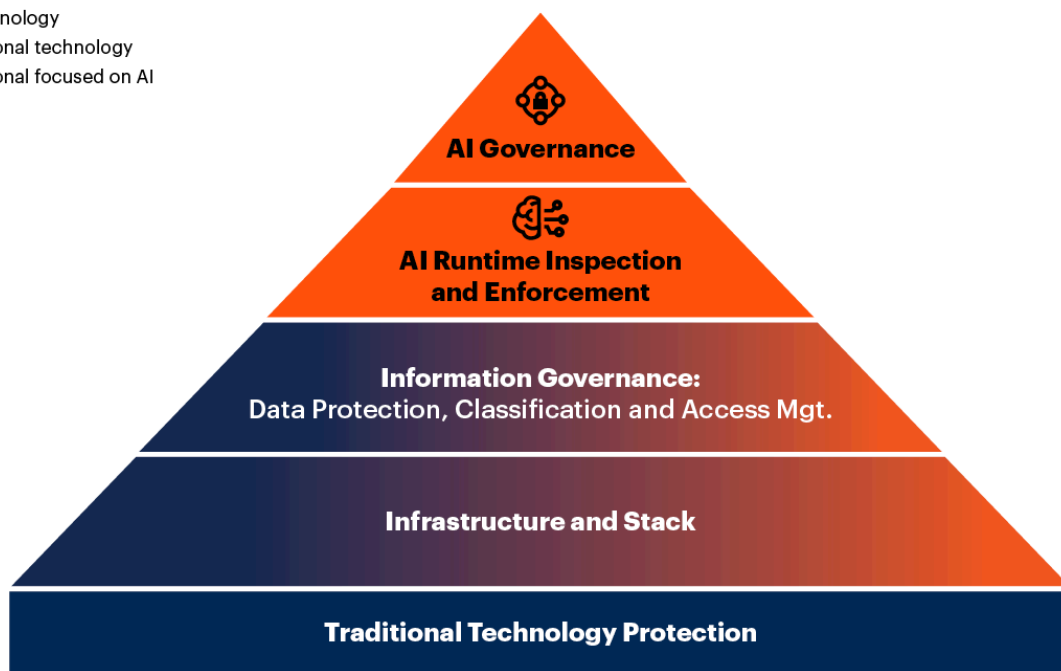
Figure 1 shows the four layers of AI TRiSM, which sit atop a fifth bottom layer consisting of traditional technology controls, such as network, endpoint and cloud security solutions.

Figure 1: AI TRiSM Technology Functions



AI TRiSM Technology Functions

- AI technology
- Traditional technology
- Traditional focused on AI



Source: Gartner
785960_C

Gartner

The top two layers of the AI TRiSM technology framework — AI governance and runtime solutions — are consolidating, in part, to form a distinctly new market segment required to manage AI risks and threats. Specifically, combining governance capabilities for AI inventory and continuous evaluations with runtime inspection and enforcement functions supports runtime analyses of AI interactions, anchored to a continuously risk scored AI inventory. This new segment will continue to be supported by enterprise solutions contained in the bottom two layers of AI TRiSM which represent traditional technology focused on AI information and workloads.

No single vendor or solution addresses all aspects of AI risks and threats at this time. Some vendors specialize in supporting security and risk mitigation, while others specialize in supporting ethical or safe practices and regulatory compliance. Even within the AI governance layer, some vendors are more focused on a wide array of governance functions, while others only provide as much as needed to effectively provide runtime controls. As noted above, these controls usually consist of AI inventory and continuous evaluation.

The AI TRiSM market is distinguished by its independence from frontier model provider solutions that also support trust

risk and security management (see Note 1). Enterprises must retain independence from any single AI model or hosting provider to ensure scalability, flexibility, cost control and trust, as AI markets rapidly mature and change.

Enterprises need to be able to pick and choose which AI tools and models best fit their use cases from a performance, cost and risk point of view. They cannot simply rely on vendors adding AI functionality to existing enterprise applications to meet all the organization's AI requirements. Enterprise AI TRiSM functions should be portable across back-end hosting providers. At the same time, enterprises should supplement, where they can, their own TRiSM services with TRiSM functions provided by hosting and frontier model providers to support a robust, layered control framework.

Over time, we expect comprehensive AI TRiSM as a service to emerge as a category of outsourced services focused on the needs of small and midsize enterprises. These services will work for a wide range of AI implementation types (see Slide 8 in [AI Technology Sandwich: A Conceptual Framework for Executing AI](#)).

What the AI TRiSM Market Guide Does Not Include

This guide does not include TRiSM solutions available from frontier model hosting providers. These companies typically provide AI filtering for socially unsafe interactions that involve violence, hatred, self-harm or inappropriate sexual interactions. Most model hosting providers support private enterprise tenants, encryption of enterprise data in transit, and data processing agreements that keep enterprises in control of their data. They are also starting to support the detection of commonly identifiable security violations, such as prompt injections.

Model hosting and frontier model providers do not typically have access to all enterprise data and cannot provide comprehensive information governance that is critical for reliable and secure AI adoption. Further, frontier model providers are not in the business of customizing AI TRiSM policy engines for enterprises' specific acceptable use policies. What they can do, however, is provide tooling for enterprises to customize and deploy those policy enforcers on their own.

Using AI can lead to existential societal risks, but the AI TRiSM Market Guide does not include solutions that directly mitigate those risks, as enterprises currently have no control over them.

Frontier model providers, AI researchers and regulators are responsible for measures that mitigate existential AI risks.

Finally, AI TRiSM is not part of the security or fraud management market that uses AI to detect malicious incidents, such as fraud, deepfakes, theft or data breaches, nor is it part of the market that leverages GenAI to make security or fraud operations more efficient.

Market Direction

Four distinct trends are shaping the AI TRiSM market and its expected future (see Figure 2). These trends have been pushed to the surface since ChatGPT democratized access to GenAI and highlighted the urgent need to control enterprise use of AI. ² Until that time, most enterprise AI development:

- Was used for first-party models developed internally by data science teams
- Required high levels of data science and AI engineering skills to work with

Thus, TRiSM took a backseat in most nonregulated industries.

Figure 2: Four Key Trends Shape AI TRiSM Market



Four Key Trends Shape the AI TRiSM Market



Source: Gartner
785960_C

Gartner

Trend 1: AI TRiSM Teams Take Shape

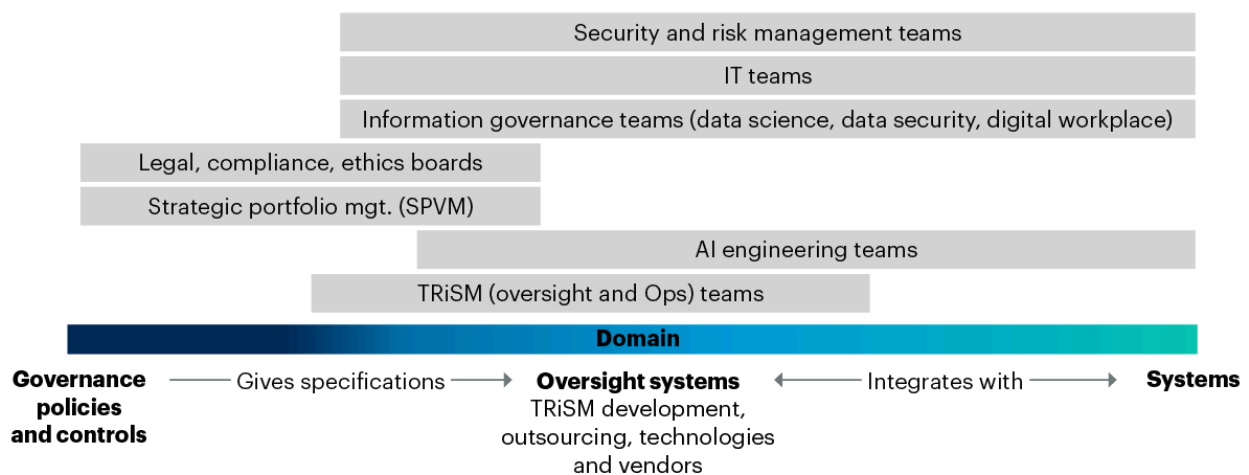
AI requires new cross-enterprise team processes and tools for effective runtime deployments. This ensures all interested parties work together in achieving unified AI outcomes that meet their different requirements. Gartner is witnessing four emerging organizational trends that will naturally lead to TRiSM oversight and operations teams over the next three years (see Figure 3):

- Security and risk management functions coalesce and integrate with relevant IT software engineering and quality assurance teams, and shift left to AI engineering to manage AI TRiSM controls.
- AI centralized or hybrid teams emerge and manage all aspects of AI, including AI TRiSM controls.
- Information governance fractures the surface, causing a renewed immediate need for a unified approach across different organizations for protecting information used by AI.
- Legal and compliance staff take demand that regulatory and legal requirements are satisfied in enterprise AI applications.

Figure 3: Teams Must Align With New AI Processes and Domains



Teams Must Align With New AI Processes and Domains



Source: Gartner

Note: There is an overlap of the teams and their relative responsibilities over the domains.

SPVM = sourcing, procurement and vendor management

785960_C

Gartner

Gartner expects AI TRiSM oversight and operations teams will report to the manager responsible for enterprise AI innovation, development and operations. Meanwhile, budget authority for AI privacy, security and risk initiatives typically sits with the office of the CTO or CIO, according to an enterprise AI survey Gartner conducted in 2024. ³

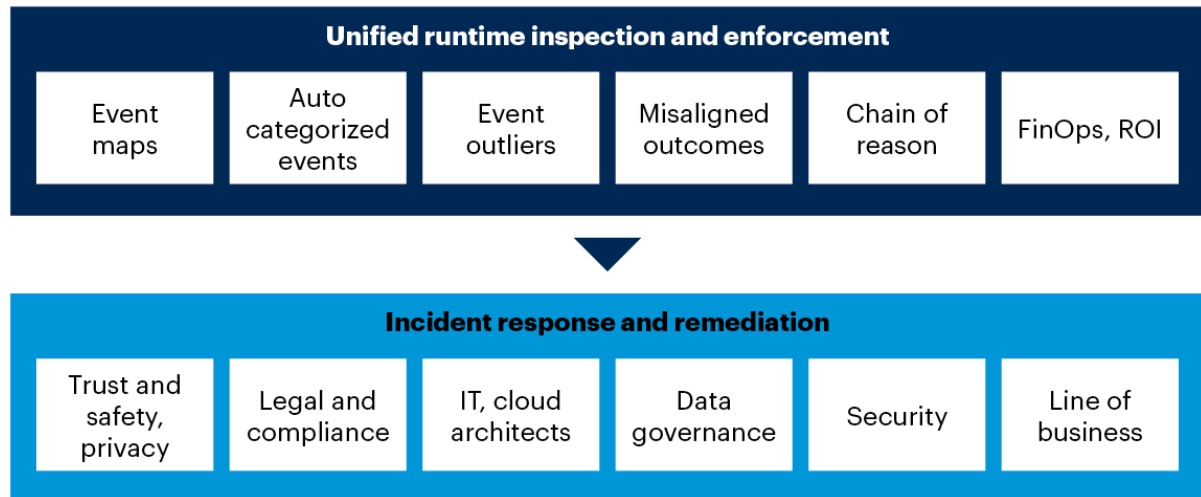
Trend 2: Unified AI Runtime Inspection and Enforcement Systems

Consistent signaling, low latency, single-pass processing and unified runtime control systems are needed to effectively process runtime inspection of AI events and enforce enterprise policies. Otherwise, siloed inspection and policy enforcement systems for the same events at runtime will result in conflicting analyses and resolutions (see Figure 4 and Note 2 for elaboration on unified AI runtime inspection and enforcement system).

Figure 4: Unified Runtime Inspection and Enforcement



Unified Runtime Inspection and Enforcement



Source: Gartner
785960_C

Gartner

It is important to:

- Execute runtime policy engines (consisting of rules and AI models) simultaneously — e.g., separate policy engines that score data compromise risk, bias risk, legal risk and security threats.
- Merge their results into a common blended risk score that guides future enterprise actions and remediations for the suspect AI interactions.

Organizations must identify the best equipped team to manage the unified runtime inspection and enforcement system, whether it is a new AI team in AI engineering, an expanded IT team or an expanded security and risk team. Regardless of who owns the unified runtime detection and enforcement system, incident response and remediation for suspect transactions that are not automatically blocked or remediated should be passed on to units and systems best equipped to handle them.

For example, under most scenarios, security events detected in the unified inspection and enforcement system would be sent to the organization's security information and event management (SIEM) team (or outsourced managed security service provider, as appropriate) for necessary followups. AI event triage will then require multiple escalation queues, based on the type of anomalies or violations encountered:

- Acceptable use violations and hallucinations would be sent to the unit best equipped to manage them — e.g., an AI incident response team.
- Prompt injections and more traditional attacks would be sent to the security and risk team.

New playbooks to triage and escalate nonsecurity AI violations, such as legal or compliance issues will be necessary.

Trend 3: AI Hosting Providers Expand TRiSM Services

Frontier model providers and other AI hosting providers are continually expanding the TRiSM services they offer enterprises as a competitive strategy for gaining more of their business. A few have already introduced tooling that enterprises can use to support their own specific acceptable use policies. Most hosting providers have implemented guardrails against use of their models for violence, self-harm, sexual content and hatred. Many have implemented tooling to detect copyright infringements and known runtime security threats, such as prompt injections.

AI TRiSM will always be a shared responsibility across users and model providers, the level of which depends on the configuration of the AI services, in-line whether they represent IaaS, PaaS or SaaS services. Organizations who host and manage their own models are solely responsible for all risk mitigations.

Trend 4: AI TRiSM Market Consolidation

Integration of AI governance with AI runtime inspection and enforcement: This has already been achieved, in part, by several vendors in this Market Guide, and we expect this trend to accelerate in 2025 and beyond. Typically, vendors providing AI runtime inspection and enforcement also provide two AI governance functions:

- AI catalogs (e.g., inventory)
- Continuous assurance and evaluations

Integrating these two governance functions with runtime controls provides the foundation for real-time analyses and policy enforcement for all enterprise AI interactions, and gives a complete view of AI risks and threats.

Land and expand strategy from multiple types of vendors: The consolidation of AI governance and runtime inspection and enforcement will integrate into services offered by established large software, security or networking vendors seeking to grow customer wallet share by supporting users in their AI journey. For instance:

- Gartner saw signs of this trend in 2024, when Cisco acquired Robust Intelligence, an early startup TRiSM vendor.
- We also witnessed cloud based network security vendors like Zscaler and Palo Alto Networks expand their capabilities into AI TRiSM.

Further:

- Some regulatory, compliance/fairness and trust and safety oriented solutions will consolidate with established governance risk and compliance (GRC) systems, many of whom are already adding modules for AI governance.
- Some AI runtime inspection and enforcement solutions will consolidate with SASE, SSE and cloud security vendors like Zscaler and Wiz, and SIEM and SOAR vendors like Exabeam or Cisco (Splunk).
- Vendors who provide outsourcing services for security and risk functions will expand to offer AI TRiSM services, targeted largely at small and midsize organizations.
- Vendors who provide enterprise AI management platforms that support an abstraction layer for AI life cycles, will increasingly integrate TRiSM capabilities into their platforms. For example, AI vendors such as [Airia](#) and [Backplain](#) already integrate TRiSM capabilities while helping users determine which back end model is the most suitable and cost-effective option for a given use case. Eventually, these AI management vendors will be acquired by larger software firms.

Information governance layer consolidation: Vendors providing data security posture management (DSPM), data security and data loss prevention (DLP) systems will continue to expand to incorporate AI-centric functionality. In particular:

- DSPM vendors will incorporate dynamic data classification and permissioning at AI runtime.

- DLP vendors will expand their functionality to support context based inspection of AI sessions and interactions, and use AI policy engines to enforce acceptable use and access controls. DLP vendors will also broaden their products to support responses and inbound information, (and not just transaction inputs and outbound information as they do today).
- Data security platform vendors will integrate security of both structured and unstructured information in their offerings.

Market Analysis

As illustrated in Figure 1, the AI TRiSM market consists of four functional technology layers. The top two are new to AI; the bottom two are traditional technology measures focused on AI:

- AI governance
- AI runtime inspection and enforcement
- Information governance
- Infrastructure and stack

Multiple capabilities reside within each functional layer. Here, we look at each layer in detail and outline those capabilities.

AI Governance

Enterprise AI governance involves several stakeholders across technical and functional teams — data science, AI engineering, legal, privacy, risk and compliance, human resources, ethics, security, audit, procurement and business teams. AI governance provides a framework to manage an organization's AI assets, instill confidence in AI performance and ensure business outcomes.

AI governance, with similar organizational pillars and focus as other governance initiatives, is driven through people and process initiatives with technology as an enabler. *This section is focused on AI governance technology.*

AI governance technology provides a unified view of the health of enterprise. AI governance technology straddles three core frameworks — responsible AI, protected and secure AI, and

AI-ready data — and provides the connective tissue across the various AI implementations within the enterprise:

- A responsible AI framework includes the principles of safety, security, accountability, transparency, fairness and ethics.
- The protected and secure AI framework includes the principles of acceptable use enforcement, alignment of AI with user intentions, and protection against malicious actors.
- An AI-ready data framework ensures access to trustworthy and protected (through access management and data classification) data for AI solutions.

In addition, AI governance technology supports assurance and audit practices for AI implementations:

- AI assurance involves various processes and activities to verify that models, agents and applications are performing according to the business, legal (e.g., contractual), regulatory (e.g., EU AI Act, General Data Protection Regulation [GDPR]) and compliance (e.g., NIST Artificial Intelligence Risk Management Framework (NIST AI RMF), ISO/IEC 42001) obligations. AI governance technology supports linkage of evidence and artifacts related to human assurance activities.
- Most AI audit requirements include system and validation traceability to ensure compliance with organizational policies and regulatory requirements.

All of this requires several core AI governance technology capabilities with extensions that support integration with operationalization platforms, most notably AI runtime inspection and enforcement solutions. See Table 1 and Note 3 for elaboration on AI governance capabilities.

AI Governance Technology Functions		
 Governance: Pre- or Postdeployment Supporting Activities		
AI visibility and traceability	Workflow (internal and third party)	Continuous assurances and evaluation (internally developed and third party)

• Catalog: Discovery, inventory and risk scoring of all AI used in organization • Automated (partly) documentation: Bill of materials, model cards, artifacts, regulatory reports, explainability • Audit trails: Audits of all state changes to AI	• Approvals of new AI; exceptions • Attestations of use • Communications with third-party providers: • RFI • Responses • Compliance • Controls	• AI security testing: Red teaming and scanning for all entities (models, applications, agents) • Risk and trust control validation: • Bias • Leakage • Trust
---	---	--

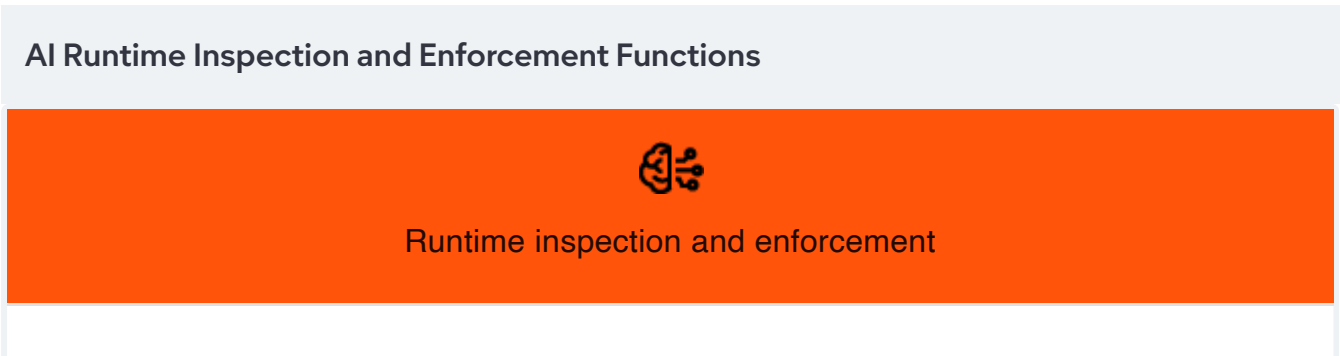
Source: Gartner (February 2025)

AI Runtime Inspection and Enforcement

AI runtime inspection and enforcement applies to models, applications and agent interactions and supports transactional alignment with organizational governance policies. Applicable connections, processes, communications, inputs and outputs are inspected for violations of policies and expected behavior. Anomalies are highlighted and either blocked, autoremediated or forwarded to humans or incident response systems for investigation, triage, response and applicable remediation.

Gartner is seeing much similarity in runtime inspection and enforcement controls applied across AI entities such as models, applications and agents. Best-of-breed runtime inspection systems will offer a modular solution architecture where users can apply applicable detections and enforcements to any AI entity — for example, model, application or agent — based on the AI use case.

Table 2 shows the different measures applied for different AI entities; these measures are further described in Note 4.



AI models	AI applications	AI agents
• Model monitoring • Detect anomalous activities and security events; block if possible and autoremediate • Data protection • Responsible AI filtering • Compliance with internal and external controls	• Context-based access control and dynamic contextual data classification • Content anomaly detection for input and output risks (safety and enterprise acceptable use); autoblock if possible • AI runtime defense (malicious and hacker attacks); autoblock if	• Align agent activity with agent intentions • Detect and flag anomalous activities and security events • Automatically block, when possible, anomalies and security events • Autoremediate, when possible, anomalies and security events • Compliance with

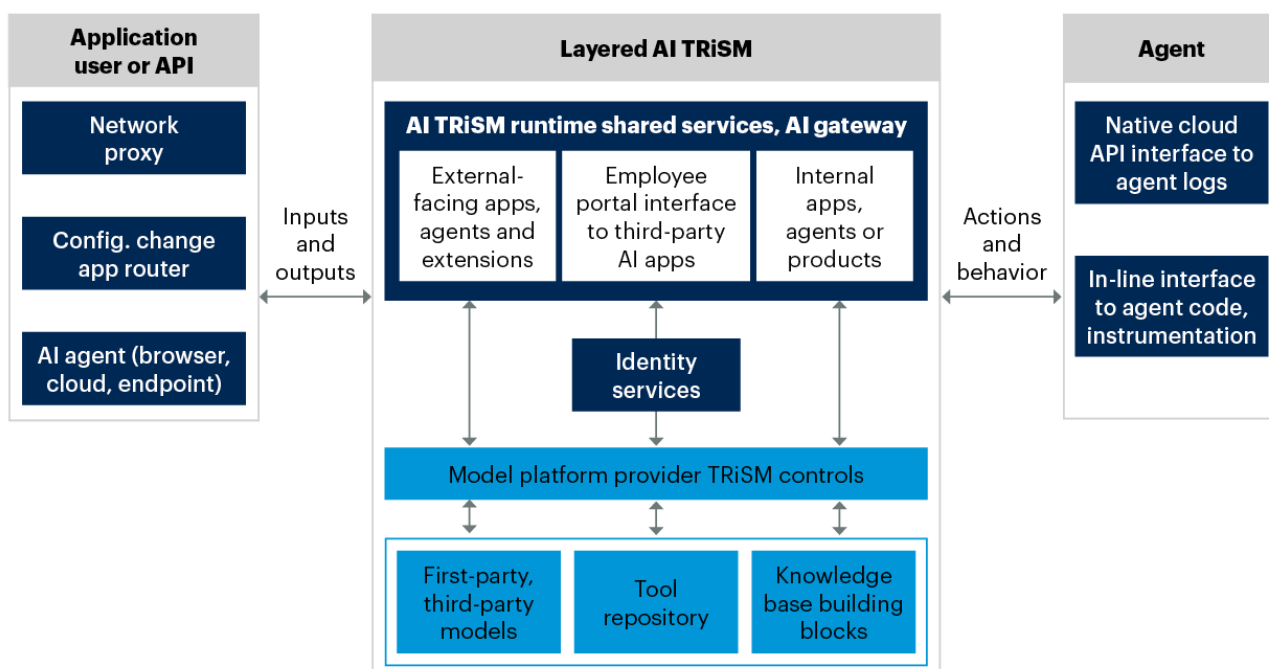
Source: Gartner (February 2025)

Figure 5 depicts a sample enterprise integration architecture for runtime inspection and enforcement for homegrown apps and agents. This diagram would look different for other types of AI, such as embedded AI in off-the-shelf applications.

Figure 5: Sample Enterprise Integration Architecture for Homegrown Apps and Agents; Runtime



Sample Enterprise Integration Architecture for Homegrown Apps and Agents; Runtime



Source: Gartner
785960_C

Gartner

Information Governance

In the context of AI TRiSM, information governance is concerned with classifying, protecting, securing and managing data throughout its life cycle. Mature information governance controls are critical for:

- Ensuring that AI only has access to relevant and correctly permissioned information
- Securing the organization's sensitive information
- Attempting to comply with relevant AI regulations, such as the EU AI Act

For more, see [Adopt 5 Key Privacy Lessons When Preparing for AI Regulations](#).

Oversharing due to inappropriate access controls and data leakage risks caused by poor information governance are not new problems. However, when GenAI tools are embedded into enterprise applications or use RAG to find organizational information, they can expose overshared data much faster and to a larger audience, and illuminate data life cycle challenges. Failing to take appropriate measures to identify and govern sensitive data throughout the organization's data estate will exacerbate such challenges, introducing unnecessary data risks and potentially culminating in a widespread data breach. This is a particular problem for well-

known horizontal AI products like Microsoft 365 Copilot, Gemini for Workspace and ChatGPT Enterprise (if integrated into productivity platforms). Unstructured data in productivity platforms like Microsoft 365 and Google Workspace has been historically poorly managed, which becomes a significant issue when AI models deployed in the infrastructure have default, widespread access to organizational data.

To guard against oversharing concerns when using GenAI and other AI tools, organizations must improve information governance holistically. The challenges are twofold:

- Most organizations and vendors have tried to address this from specific siloes — e.g., security, data protection, identity, or compliance — rather than improving information governance by bringing together the siloed groups for a consolidated view and coordination actions.
- Information governance often becomes an afterthought, and can be perpetually delayed in deference to other projects that are deemed more pressing (see Note 5).⁴

For many organizations, weak information governance is emerging as the major obstacle to wider GenAI rollouts and presents an opportunity for third-party vendors specializing in this space.

To improve information governance, organizations need to embrace a multidisciplinary team consisting of stakeholders from various teams, including digital workplace, data and analytics (D&A), data security, compliance, and identity and access management (IAM). Specifically, this multidisciplinary team must focus on:

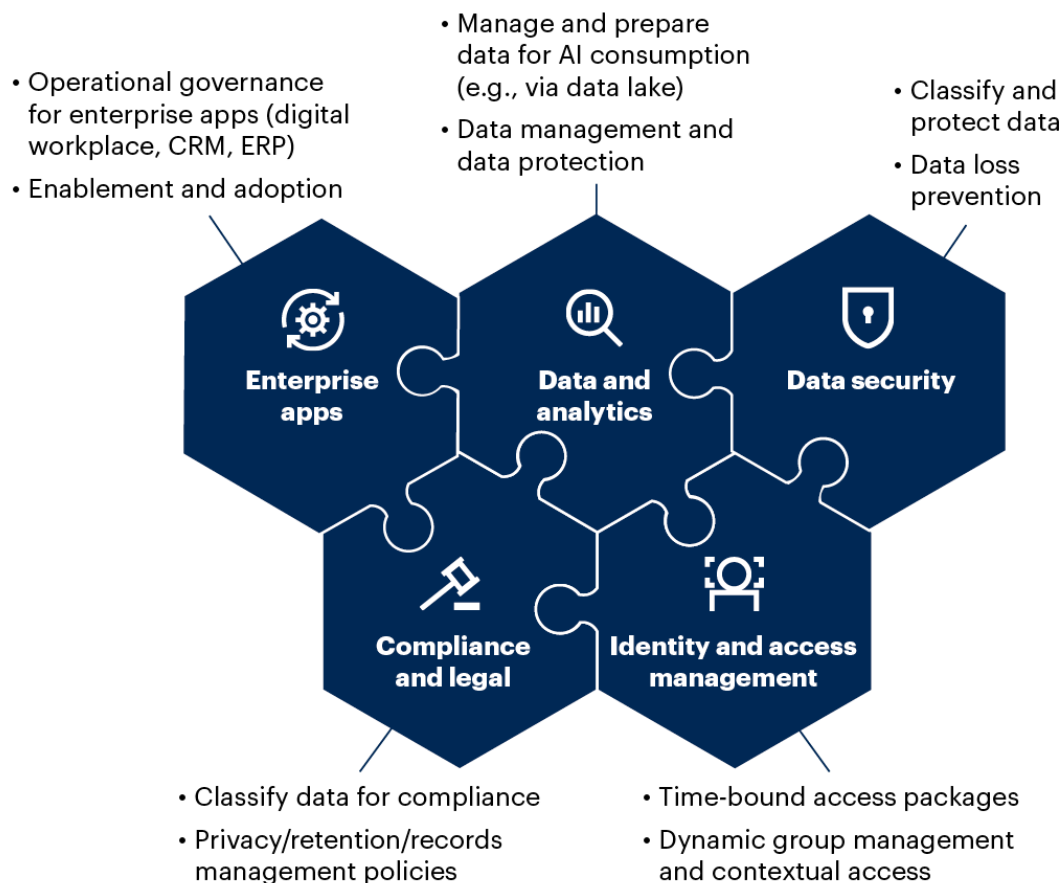
- Bridging the gaps between the various representative teams to ensure that the organization's security, usability, privacy and scale requirements can be met (see Figure 6). Do this by, for example, facilitating predeployment workshops that are not specific to any security discipline, but to the organization's strategic business outcomes.
- Processing, classifying, protecting and securing sensitive information in accord with data purpose, organizational objectives, and business needs.

- Establishing data retention policies to manage redundant, obsolete or tertiary content (ROT).
- Switching to a policy of purpose-based access control (PBAC) to cut down on “permissions sprawl.”
- Building scalable information architectures in productivity platforms (like Microsoft 365 and Google Workspace) to make the above simple to manage and easy for users to understand.

Figure 6: Multidisciplinary Approach Required for Information Governance



Multidisciplinary Approach Required for Information Governance



Source: Gartner
785960_C

Gartner.

Currently, this technology layer is still fragmented with vendors that usually focus on one or two of the required components for comprehensive information governance. For example:

- DSPM, DLP and DAG vendors focus on data security.

- Identity governance vendors focus on permission management.
- Compliance vendors focus on retention and data life cycle.

Privacy-enhancing technologies (PET) are a collection of tools that help protect data in use, such as during model training, fine tuning, inference on obfuscated data, or federated model execution across entities that might not want to share data while information exchange remains a deliberate purpose. PETs can be applied in several layers of the AI TRiSM stack, such as at the compute level (for example, confidential computing), in information governance (for example, using consistently and robustly altered data across life cycle stages using synthetic data or forms of homomorphic encryption such as Enveil), in performance monitoring (for example, using synthetic data), or obfuscation in inference, RAG and fine-tuning (for example, Protopia) (see Note 6 on PET).

As such, while many products can be used to reduce information governance risk, you will be unlikely to find a single product that offers a comprehensive solution. See Note 7 for description of key information governance functions.

Infrastructure and Stack

Infrastructure and stack refers to the hardware and software environments that support AI workloads, special-purpose tools and libraries used for development and maintenance of AI applications, the data processing layer and the deployment environment including the network infrastructure. Gartner is observing solutions in this layer being applied to AI workloads, thereby fortifying the protection of AI data and processes. Solutions in this category include API key management and confidential computing to protect workloads in trusted execution environments.

Representative Vendors

The vendors listed in this Market Guide do not imply an exhaustive list. This section is intended to provide more understanding of the market and its offerings.

Vendor Selection

Vendors are diverse and fragmented in their AI TRiSM capabilities. Inspect the products under consideration for the detailed functionalities included in each of the core AI TRiSM

capabilities listed in the Market Analysis section.

AI Governance and Runtime Inspection and Enforcement

AI governance and AI runtime inspection and enforcement solutions are new to AI and constitute the top two layers of the AI TRiSM technology framework. These two layers of AI are integrating in part to form a distinct new segment of the AI TRiSM market.

At this time, vendors providing AI governance and runtime controls differ considerably in their capabilities, and Gartner observes two primary categories of vendors in this space (with some vendors crossing both categories in part):

- **Policy governance:** These are vendors who have primarily and historically targeted GRC buyers by providing full AI documentation, regulatory compliance monitoring and a wide range of AI governance functions as defined in Table 1 above. Many of these vendors have also branched into runtime controls that validate enterprise policies especially around compliance are being enforced.
- **Operational governance:** These are vendors who target buyers responsible for ensuring a wide range of risk and security threats are mitigated during AI operations. These vendors offer a narrower range of AI governance functions (i.e., AI inventory and continuous assurances/evaluations) and support runtime analyses of AI interactions anchored to a continuously risk scored AI inventory.

Table 3 lists sample vendors that fall into both of these categories, where the lines are starting to blur as vendors expand their feature set to satisfy user demands.

Representative Vendors: AI Governance and Runtime Inspection and Enforcement

Company	Product	Headquarters	Year founded	Differentiating capability

Source: Gartner (February 2025)

Information governance vendors included in the table below represent sample vendors who focus their product efforts on protecting information for AI usage. As noted, sample PET vendors are listed in the referenced research in Table 4.

[/www.gartner.com/doc/reprints?id=1-2KCRQRQ4&ct=250224&st=sb](http://www.gartner.com/doc/reprints?id=1-2KCRQRQ4&ct=250224&st=sb)

AvePoint	The AvePoint Confidence Platform	Jersey City, NJ	2001	The AvePoint Confidence Platform is the single pane of glass
----------	--	--------------------	------	--

Source: Gartner (February 2025)

Market Recommendations

AI leaders must work with stakeholders across the organization to manage AI trust, risk and security. They should establish an organizational structure for creating and updating AI governance policies and for evaluating and implementing AI TRiSM technologies that enable and enforce these policies. This will help the organization systematically ensure safer, and more reliable, trustworthy and secure AI use.

Once organized, AI leaders and their business partners should:

- Discover and inventory all AI used in the organization, leveraging the capabilities of vendors who help with this, as defined in the AI governance section of this guide.
- Revisit and implement data classification, protection and access management across all enterprise information — both structured and unstructured — that can potentially be utilized by AI. This will necessitate collaboration across different teams involved in information governance.
- Work with legal and compliance experts to contractually define accountability and responsibility for unacceptable AI use or behavior in third-party embedded AI applications.
- Evaluate and implement layered AI TRiSM technology to continuously support and enforce policies across all AI entities in use. This includes enterprise-owned AI TRiSM services, along with controls offered by frontier model providers. Evaluate frontier model hosting provider capabilities from vendors such as Microsoft, OpenAI, Amazon Web Services (AWS), and Google and apply them where appropriate but do not solely rely on them:
 - Enterprises must have controls that enforce their own enterprise specific AI TRiSM policies and simultaneously maintain provider independence. This way enterprises can use the most effective hosted AI for their use cases, whether they are large language models, small language models, open source models or any other type of AI tooling.

- Enterprises that don't have sufficient resources to assemble these services on their own should rely more heavily on third party providers of embedded AI solutions for their TRiSM controls.
 - Ensure independence from any single AI model provider to enable future flexibility as AI evolves and more advanced and cost-effective AI models become available. TRiSM services help support that independence.
 - Monitor the vendor landscape continually for expected consolidation and for emerging TRiSM outsourced services that reduce the burden of system integration and maintenance.
-

⊕ Evidence

Note 1: Frontier Model Provider Definition

A frontier model provider refers to a company that develops and operates the most advanced AI models that are pushing the boundaries of what AI can achieve — e.g., multimodal AI or artificial general intelligence. This includes companies, such as OpenAI, Google, Microsoft, Anthropic and others. The models they are developing or helping develop are much more powerful than the AI models currently on the market.

Note 2: Unified Runtime Inspection and Enforcement System

Different units and participants will have different interests in the trust, risk and security of AI interactions. For example:

- The trust and safety personnel are interested in identifying safety issues so that customer-facing chatbots don't lead to customer self-harming activities.
- Line-of-business units want to ensure customer-facing AI applications conform to policies they establish, so that for example customers are not offered discounts that do not exist.
- The security team is intent on blocking malicious code injected into agent traffic that hijacks enterprise assets for malevolent purposes.

- The information governance team seeks to block potential data compromise when accessed information is not properly classified or permissioned.
- The legal team wants to block transmission of enterprise intellectual property to external AI models and conversely they want to block receipt of others' IP that the enterprise is not entitled to access.

In sum, the same AI transaction may be high risk with one lens but low risk with another so a unified single pass processing system is needed to make the decision to alert or block based on multiple considerations.

In the scenario above transactions that are not automatically blocked will be passed on to different incident and response teams if they exist. For example:

- Harmful customer chatbot interactions detected at runtime should be sent to a unit that knows how to respond to trust and safety issues.
- Security events should be forwarded to security incident response systems.
- Data compliance violations should be forwarded to information governance teams who can respond and remediate to prevent similar future violations.
- Legal and compliance issues should be sent to a compliance staff that can analyze the violation and respond accordingly.

Note 3: AI Governance Technology Capabilities

AI visibility and traceability:

- **AI catalog:** The catalog provides a comprehensive inventory of all AI entities used in the organization, e.g., models, agents, applications in various configurations ranging from embedded AI in off-the-shelf third-party applications to bring your own AI to retrieval augmented generation (RAG) systems and to first party models. The catalog risk scores these entities, and provides a structured way to track and access them. The catalog also stores the metadata of the AI entities — for example a model card and its run time metrics. This enables discovery by business and technical users. The catalog is agnostic to the operationalization platform.
- **AI documentation:** This includes model cards, bill of materials and regulatory reports

- **AI ownership:** The platform should identify the ownership of various artifacts and activities, for example, the model owner or risk assessment owner.
- **AI risk and regulations:** The usage of AI systems raises several risks and requires compliance with various industry and government regulations. The AI governance platform includes the risk assessment for the use-case artifacts (models, applications and agents), along with the constraints and requirements to confirm regulatory compliance. This may in the form of validations, documentation, or reports.
- **AI data mapping and lineage:** Data is used to train or fine-tune AI models. Data is also used to provide context for user queries in a RAG system, or to feed AI agent systems. AI governance technology functions should map data used by various AI entities and track the usage over time. It may interface with data governance platforms to include data lineage, classification, ownership and data observability information.
- **AI audit:** The AI use-case implementation involves several components, including data, prompts, models, applications and agents. The use-case output is used for decision making. The AI governance platform supports the audit of the complete life cycle through lineage, versioning and traceability.

Workflow:

- **Approvals:** All new AI systems need to be approved for specific usage. This may include reviews and approvals from AI governance boards, procurement teams, architecture review boards, security team and risk teams.
- **Attestation and validation:** New AI systems or updates may require an ongoing attestation and validation as part of the workflow for AI system implementation. The workflow includes internal and external AI along with third party embedded AI. Some of the examples are machine learning (ML) models developed internally, or external LLMs accessed through APIs integrated within a RAG solution or a conversational chatbot, or a third-party HR solution with embedded AI models.
- **Communication with third-party providers:** AI systems are not limited to internal development. Business function focused solutions or on-premises, from third-party system or solution providers involve providing a RFI, assessment of responses, compliance and controls requirements. Some AI governance systems are providing structured processes for communications between buyers and their suppliers. In theory, supplier responses can

be shared across different customers, and validated when systems are deployed through systematic and continuous testing.

Continuous assurances and evaluation:

- **AI security testing:** Security testing of the various AI entities used to access, develop, run or maintain AI such as notebooks, endpoints, models, agents, applications, and pipelines. These tests aim to protect enterprise assets against adversarial attacks. Continuous automated testing and code scans are essential parts of a red teaming program for a trustworthy and secure AI implementation.
- **Risk and control validation:** This determines deviations or anomalies related to bias, leakage, trust and use-case alignment.
- **Posture management:** The metadata across the end-to-end life cycle and operations provide a consolidated view of the business requirements, design time metrics and runtime metrics. Integration with runtime inspection and enforcement systems can turn static posture management into real-time posture management.
- **Compliance reporting:** Internal and external reporting requirements to support compliance with various risk and regulatory frameworks.

Note 4: AI Runtime Inspection and Enforcement Layer Components

TRISM controls for AI models fall under the responsibility of model builders or owners.

Organizations using a third-party model can implement some of these controls, treating the model as a third-party component:

- **Model monitoring:** Sitting between the application stack and the model, model monitoring inspects all inputs sent to the model and raw outputs of the models, before validation and modification by the application stack. This includes model outputs that become inputs to other models, so the chain of model events are monitored across models.
- **Anomalous activity and security event detection; block if possible and autoremediate:** Detect and flag anomalous activities and security events. Auto block high risk activities and events, when certain. Respond and autoremediate for similar future anomalous or security events so that the same issues do not reoccur.

- **Data protection:** At runtime, the scope of model data protection is to protect the model itself, primarily preventing retrieval of training data as well as protecting inference data at runtime.
- **Responsible AI filtering:** Ensures that the model outputs comply with model owner defined AI safety, ethics and fairness measures, and remains in scope for its intended use, avoiding harmful or toxic outputs.
- **Compliance:** Map model controls and deployment characteristics with global and local regulations requirements, often in the form of a compliance rate.

AI applications:

- **Runtime data protection:** Enforce least privilege access/role-based access control to ensure that retrieved content is limited to what the user of the application can legitimately access. Also performs dynamic (real-time) content classification so the access controls can be properly applied to data that is sent or retrieved by users or machine processes.
- **Content anomaly detection for inputs and outputs risks:**
 - **Inputs:** Detects anomalies in inputs as defined by enterprise policies for example to detect violations in acceptable use or in acceptable data inputs.
 - **Outputs:** detects anomalies in outputs to similarly support enterprise policies. This includes detection of:
 - Outputs that violate enterprise acceptable use policies
 - Hallucinations
 - Inaccurate, illegal, copyright-infringing and otherwise unwanted outputs that compromise enterprise decision making
- **AI runtime defense:** Detect malicious activities, such as prompt injections, exploitation of software component vulnerabilities to gain privileged access, or attempt to bypass existing guardrails (e.g., jailbreaks). Optionally can prevent inputs to be transmitted to the model, or block the response before it reaches the user, and supplement it with a warning message.
- **Intelligent data obfuscation and redaction:** Apply privacy enhancing technologies, such as data masking or data redaction automatically on input and outputs to preserve user and

enterprise data privacy.

- **Compliance:** Monitor in real time applications inputs and outputs, and map them to global and local compliance requirements to detect potential compliance violations.

AI agents:

- **Agent alignment:** Ensure that the agent activities are in scope with agent intentions.
- **Anomaly detection:** Detects and flag anomalous agent activities and security events. Autoblock high risk activities and events, when certain.
- **Autoremediation:** Respond and autoremediate for similar future anomalous or security events so that the same issues do not reoccur.
- **Compliance:** Monitor in real time agent processes and activities inputs and map them to global and local compliance requirements to detect potential compliance violations.

Note 5: Information Governance Survey Data

Recent Gartner survey data indicate that for some organizations, security teams lead many of the downstream objectives within information governance (like deciding on DLP technology for minimizing data loss), but have less responsibility for leading upstream efforts (like establishing data discovery and determining data classification schemas).¹ Specifically, 74% of survey participants indicate that security leads the decision making for DLP technology, but indicate that security is less likely to lead the function in other critical information governance efforts, like undertaking data discovery or determining their organization's data classification schema. Respondents indicated that security is only "consulted as needed" in undertaking data discovery efforts (lead decision maker for 48% of respondents) and in deciding the data classification schema (lead decision maker for 54% of respondents). If security is not the leader of the activity, the leader in undertaking data discovery becomes either data and analytics (26%), IT (18%) or a cross-functional group (17%). If security is not the leader of the activity, the lead decision maker for the data classification schema becomes either a cross-functional group (29%), data and analytics (20%), or the data/asset owner (15%). It should be noted that the percentage (where security is not the leader of the data discovery effort or the lead decision maker in data classification) is "percentage of respondents that said that cybersecurity isn't leading this activity" and not the full N size of respondents.

This survey points to the fact that Information governance functions in most organizations are fractured, and AI has surfaced the difficulties protecting information in siloed units. Information governance starts with data discovery and classification but ownership for this important function often falls through the cracks between compliance and data security. Information governance also includes identity access management, data security and digital workplace information protection. Organizations typically try to meet separate objectives defined by a data security or compliance team, rather than those that promote all aspects across the life cycle of enterprise information.

Note 6: Privacy-Enhancing Technology

Privacy-enhancing technologies (PETs) are the collection of tools to fill that glaring gap in many day-to-day use cases. Though PET is an umbrella term, there is a benefit from using these techniques in specific combinations. They can be applied on different levels: on the data itself, on the software or compute level, and even on the hardware level.

Data can be transformed, either in place (homomorphic encryption), in a new repository (synthetic data) or on the fly (differential privacy). On the software or compute level, safeguards can be added through federated machine learning, zero-knowledge proofs (ZKPs) and secure multiparty computation (SMPC). Forms of homomorphic encryption can also be used to analyze or share data while either the data or the analytics algorithm itself, or both, remain in an encrypted state in use. Finally, on the hardware level, there are opportunities to let operations take place in isolated environments or secure enclaves, using confidential computing. Ideally, selecting techniques on multiple levels in combination provides more robust protection.

Sample vendors include CryptoLab, Duality, Enveil, IBM, Inpher, Intel, Snowflake (LeapYear), Lorica Cybersecurity, Informatica (Privitar), Tonic.ai, TripleBlind, Protopia.

See the profile on PET in [Emerging Tech Impact Radar: Edge Artificial Intelligence](#).

Note 7: Key Information Governance Functions

data security	Controls designed to safeguard digital information from risks including unauthorized access, disclosure, modification or noncompliance. Data security controls are also used to protect the confidentiality, integrity and availability of an organization's data.
DSPM	Data security posture management; discovers and classifies sensitive information and determines the risk associated with such data
DAG	Data access governance; concerned with identification of permissions to sensitive data and takes action to enforce appropriate permissions.
DLP	Data loss prevention; applies controls to ensure that sensitive data does not move to unauthorized destinations, typically outside the organizational perimeter.
information governance	Controls to ensure that data is secured and governed throughout its life cycle.

Disclaimer

This research may not reflect current market conditions after 18 August 2026 for this Market Guide.

Unlock more AI value

With unmatched, proprietary AI insights from Gartner

[Explore the AI Hub](#) ↗

© 2026 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.

[About Us](#) [Search Careers](#) [Newsroom](#) [Policies](#) [Site Index](#) [Featured Insights](#)
[Contact Us](#) [Send Feedback](#)

Gartner

© 2026 Gartner, Inc. and/or its Affiliates. All Rights Reserved.