Thomson Reuters™

AI Everywhere, Secrets at Risk: A Modern Framework for Trade Secret Protection

by Baldo Vinti and Wai Choy, Proskauer Rose LLP, with Practical Law Intellectual Property & Technology

Status: **Law stated as of 26 Mar 2026** | Jurisdiction: **United States**

This document is published by Practical Law and can be found at: **content.next.westlaw.com/w-049-5500**
Request a free trial and demonstration at: **tr.com/practicallaw-home**

An Article that examines the significant risks that generative artificial intelligence (GenAI) poses to trade secrets as it becomes increasingly integrated into core business functions and provides practical mitigation strategies. This Article explains how the use of third-party, cloud-based GenAI solutions can lead to the inadvertent loss of trade secret protection through user inputs, file uploads, and the potential use of confidential information for model training. It outlines how the legal requirement to take "reasonable measures" to maintain secrecy must now account for the unique challenges of AI-enabled workflows. This Article also explores the expanding risk surface created by the shift from prompt-based tools to more autonomous AI agents capable of multi-step task execution without direct human supervision. To address these challenges, this Article presents a practical framework for deploying GenAI at scale while preserving trade secret rights, offering guidance on governance and policy, vendor due diligence, contract negotiation, technical controls, and incident response readiness.

Generative AI (GenAI), a subset of AI that creates new text, visual or auditory content based on patterns learned from its training datasets, as well as inputs and reference data provided by the user, is increasingly embedded in core business functions. It is broadly applicable across a wide range of use cases, from drafting contracts and analyzing financial statements to writing software code, accelerating drug discovery, optimizing manufacturing operations, creating summary notes of meetings, and beyond.

Trade secret protection matters more than ever in the AI era because the assets that create real competitive advantage, such as confidential data, playbooks, methods, and code, can be exposed through routine model use if appropriate precautions are not implemented. When an organization selects enterprise-grade GenAI tools and uses them in accordance with well-calibrated internal GenAI and trade secret policies, safeguards and training, those tools can be safely used with trade secrets and aid in their development. Failure in any of those areas, however, can result in value-destroying loss of trade secret protection and revelation to third parties. Further, any involvement of autonomous AI agents, which are capable of executing GenAI and other functions without human supervision or intervention, can create additional complications.

This article explains the risks that GenAI can pose to trade secrets and sets out a practical framework for protecting trade secrets across GenAI workflows. The objective is to equip organizations to deploy GenAI at scale while preserving the secrecy protections that underpin trade secret rights.

For an overview of key legal issues relating to GenAI, see Practice Note, Key Legal Issues in Using Generative AI: Overview (US). For a collection of resources concerning AI legal issues, see AI Toolkit (US).

For more information on trade secret protection, see Practice Note, Protection of Employers' Trade Secrets and Confidential Information.

On-Prem vs. Vendor-Hosted

Some organizations host GenAI models entirely "on-prem" (installed on a company's own physical hardware located within their own facilities, such that no data leaves its walled environment for processing) or in a private cloud environment, thus avoiding many (although not all) of the potential risks associated with vendor-hosted GenAI tools. However, most are using cloud-based GenAI solutions in which the third-party service provider hosts the GenAI model and prompts and inputs are sent to the service provider's servers, where they are processed to generate outputs. The prompts, inputs and outputs may be retained on those servers. These vendor-hosted offerings are often favored, as they provide access to the latest models and features in a rapidly evolving GenAI market. For a model AI SaaS Agreement, see Standard Document, AI SaaS Agreement.

GenAI-Specific Implications for Trade Secrets

In addition to the typical SaaS issues, as employees prompt GenAI models, upload files, and connect internal systems to vendor-hosted GenAI tools companies must carefully consider technical, operational, and practical risks that are unique to, or heightened by, GenAI. If not appropriately managed through properly calibrated policies and guardrails, such use can result in the inadvertent loss of trade secret protection.

The technical potential for GenAI models to train on additional data, and GenAI's distinguishing feature (creating new content derived in part from its training data in response to a user's prompt), raises a risk that does not arise with other technologies: Absent clear contractual restrictions and technical safeguards, customer inputs and outputs may be retained and used for model training or fine-tuning, – creating a risk of the model being able to recreate those inputs and outputs (or parts thereof) in responses to prompts by third parties. Where those inputs and outputs include trade secrets, this could result in disclosure to a third party without confidentiality protections, which is inconsistent with the preservation of trade secret status.

The GenAI Landscape: A Broad Spectrum of Confidentiality and Security

Not all GenAI tools, or the contractual terms and conditions that govern them, are created equal. On one end of the spectrum are enterprise-grade, secure, business-ready tools governed by contracts that robustly protect the confidentiality and security of the user's inputs and outputs and the user's proprietary rights in them. On the other end are free public tools governed by terms that entirely lack confidentiality and security commitments and expressly permit the service provider to use all inputs and outputs to train AI models and disclose them to third parties – a framework that is incongruent with the fundamental tenets of trade secret protection.

In addition, even where a GenAI tool does not use any inputs or outputs for model training, the ways in which GenAI tools are configured by organizations or used in practice can create heightened risks of forfeiting trade secret status. For example, settings that store trade secret outputs in shared areas of company systems that are accessible by personnel outside the "need-to-know" circle could jeopardize their trade secret status and increase the risk of disclosure outside the company.

In this context, the risk rarely looks like theft. It comes in the form of ordinary use, such as using GenAI tools to summarize a clinical protocol, debug proprietary software source code, analyze internal transaction data, or generate an internal report or summary. Yet once confidential information is uploaded into vendor-hosted GenAI tools, retained in logs or telemetry or shared through connectors, secrecy may be lost if appropriate safeguards are not implemented. Because trade secret protection depends on maintaining reasonable secrecy measures, uncontrolled disclosure can be legally fatal.

Trade Secret Law Fundamentals

Well-kept trade secrets deliver durable competitive advantages, such as faster time to market, superior pricing and margins, and differentiated customer experiences, products and services. The market value of secrecy is reflected in the substantial damages courts have awarded for misappropriation, as well as injunctions that can reshape competition.

Under the federal Defend Trade Secrets Act (DTSA) (18 U.S.C. §§ 1836) and widely-adopted state laws derived from and generally harmonious with the Uniform Trade Secrets Act (UTSA), information qualifies for trade secret protection if both:

- The information derives independent economic value from not being generally known or readily ascertainable.
- The owner takes reasonable measures to keep it secret.

(18 U.S.C. §§ 1839(3).) The DTSA explicitly does not preempt these state laws, and trade secret owners frequently assert both state UTSA and federal DTSA causes of action where there is a basis for federal jurisdiction (18 U.S.C. § 1838).

The owner may then have remedies against a third party for trade secret misappropriation if the third party acquires, uses, or discloses such trade secret through improper means or in breach of a duty (18 U.S.C. §§ 1839(5)). DTSA remedies can include injunctive relief, damages, unjust enrichment, and enhanced damages for willful and malicious misappropriation (18 U.S.C. § 1836). As there is no statutory definition of willful or malicious under the DTSA or UTSA, when determining whether a plaintiff is entitled to enhanced damages courts consider factors such as the degree of intent, the defendant's state of mind, the extent of the harm, the duration of misuse, efforts to conceal misuse, prior conduct of similar misappropriations, and deterrence (see, for example, *Propel Fuels, Inc. v. Phillips 66 Co.*, No. 22-CV-007197 (Cal. Super. Ct. July 30, 2025) (awarding plaintiff $604.9 million in damages, including $195 million in exemplary damages, after it found that the defendant willfully misappropriated the plaintiff's trade secrets after conducting due diligence in connection with a potential acquisition of the plaintiff)).

Unlike patents, trade secrets do not come with a presumption of validity, and a trade secret owner bears the burden to show that the relevant information meets the statutory definition of trade secret (see *Synopsys, Inc v. Risk Based Sec., Inc.*, 70 F.4th 759, 769 (4th Cir. 2023)). In practice, the outcome of a trade secret misappropriation claim often hinges on whether the owner can demonstrate consistent, reasonable secrecy measures tailored to the nature of the information and how it is used. In GenAI-enabled environments, those "reasonable measures" must consider the unique risks of GenAI (see Reasonable Measures for GenAI: A Practical Framework).

For more on trade secret misappropriation claims under the DTSA, see Practice Note, Employment Litigation: DTSA Claims.

Where Trade Secrets Can be Exposed in AI-Enabled Workflows

Trade secrets could potentially be exposed through ordinary GenAI usage channels, including prompts and file uploads, connectors to email, chat, and document repositories, browser plug-ins, unmanaged sandboxes, evaluation platforms, retained logs and telemetry, and externally shared outputs. Across sectors, exposure tends to concentrate in high-impact workflows where sensitive information is integral.

These assets are valuable and fragile. An effective program begins by identifying what constitutes a trade secret, mapping how those assets move through AI workflows, and understanding where disclosure can occur.

AI Agents, Autonomy, and Enterprise Workflows

Enterprise AI is rapidly evolving from prompt-based chat tools to agentic systems capable of multi-step task execution. Traditional GenAI tools respond to discrete user inputs. By contrast, agentic systems can plan tasks, invoke external tools, access internal systems, retrieve documents, execute code, and trigger downstream actions as part of a coordinated workflow, without step-by-step human supervision. As AI agents become increasingly autonomous and interconnected through emerging open standards, the "reasonable measures" inquiry extends beyond human conduct to machine behavior.

This heightens concerns around "prompt injection," a type of exploit in which an attacker causes an AI system or agent to follow malicious or unintended instructions – whether through direct prompts or instructions embedded in untrusted content it processes – potentially resulting in adverse effects such as unauthorized data disclosure, data exfiltration, deletion, or other unintended actions.

Not all agentic systems, however, are fully autonomous. Some operate reactively – executing multi-step workflows only when triggered by a user. Others are configured with limited or persistent

autonomy, enabling them to monitor inputs, initiate actions, or carry out predefined objectives without step-by-step human supervision. The trade secret implications vary depending on where a deployment falls along this autonomy spectrum.

Leading model providers now support this architecture. Advanced GenAI systems can call external functions, interact with structured environments, retrieve contextual data, and execute multi-step workflows through API integrations. These capabilities enable the creation of agents that operate across enterprise systems rather than within a single chat interface.

The agentic AI landscape has matured significantly. In late 2025, OpenAI, Anthropic, and Block helped launch the Agentic AI Foundation under the Linux Foundation, contributing open interoperability initiatives including Anthropic's Model Context Protocol (MCP) and OpenAI's AGENTS.md specification, with support from Google, Microsoft, and Amazon Web Services (AWS). These developments reflect an industry-wide effort to standardize how agents communicate, access tools, and coordinate actions, signaling that multi-agent architectures are moving from experimental prototypes toward production-oriented enterprise infrastructure.

At the same time, self-hosted AI assistants that integrate directly into existing communication channels are becoming increasingly available. They can monitor incoming messages, execute predefined workflows, or automatically generate outputs and send communications when certain triggers are met. For example, OpenClaw (formerly Moltbot/ClawdBot) is designed to be installed and run directly on a user's computer and can be allowed to control messaging platforms such as Slack, WhatsApp, Telegram, Discord, and Google Chat. Rather than requiring users to log into a separate AI interface, the assistant can operate within familiar channels, responding to questions, retrieving documents, summarizing threads, executing tasks, and interacting with internal systems through configured connectors.

The breadth of system access that this category of AI tool can obtain raises governance concerns, including surveillance risk, data aggregation, connector overreach, and API misuse.

In practice, an agent may be configured to:

- Monitor incoming messages and respond automatically.
- Retrieve internal documents from multiple repositories.
- Query structured datasets.
- Generate reports or analyses.
- Push outputs back into collaboration platforms or enterprise systems.
- Send external communications or otherwise exfiltrate data.

In effect, the agent becomes a software actor embedded inside the organization's information architecture.

This shift materially expands the trade secret risk surface. Unlike a single prompt entered into a chat window, an agent – particularly one with workflow or persistent autonomy – may retrieve, synthesize, and redistribute information from numerous documents, databases, and communication threads during the course of a task. If connector permissions (access rights) are overbroad, API credentials are insufficiently scoped, or retention policies misaligned, sensitive information can be compromised and propagated at machine speed.

Self-hosted deployment does not eliminate this risk; it relocates it. For example, in early 2026, the OpenClaw framework disclosed that 341 malicious "skills" (modular plugin or extension designed to give the AI agent new capabilities) had been identified in its ClawHub marketplace, 335 of which stemmed from a single supply chain attack. Security researchers warned users to isolate such bots from personal files, email, and business data, underscoring that self-hosted infrastructure does not eliminate risk but relocates it to internal configuration and security controls. When organizations deploy agent frameworks internally, secrecy protections depend on infrastructure configuration, encryption, network segmentation, credential scoping, logging discipline, and ongoing access governance. Messaging platform integrations may expose confidential information through archived threads, inherited permissions, or third-party retention settings.

From a trade secret perspective, the key evolution is this: the "reasonable measures" inquiry increasingly extends beyond regulating human inputs into AI systems to governing machine-authorized access and behavior. It is no longer sufficient to control what employees type into a model. Organizations must

also define what AI systems are permitted to retrieve, combine, store, monitor, and transmit, particularly when those systems are granted workflow-level or persistent autonomy.

For more on the legal risks and challenges arising from agentic AI use, see Article, Agentic AI: Greater Capabilities and Enhanced Risks.

Reasonable Measures for GenAI: A Practical Framework

For trade secret assets, such as pricing models, customer lists, proprietary datasets, source code, and operational playbooks, trade secret protection hinges on a single question: Did the company take reasonable measures to keep the information secret?

To maintain trade secret protection, organizations must implement "reasonable measures" that account for the unique considerations raised by GenAI. The strongest trade secret programs combine clear governance, disciplined contracting, technical controls embedded directly into GenAI workflows and systems administration, and auditable processes reinforced by training and culture. The touchstone is operational consistency to write the rules, implement the controls, and log the evidence.

The strongest approach is layered and practical, including:

- Adopting and internal GenAI policy, particularly regarding crown-jewel assets (see Governance and Policy).
- Tightening policies and contracts, including disciplining vendor access (see GenAI Vendor and Model Provider Due Diligence and Contracts).
- Reinforcing personnel training (see Training).
- Embedding technical controls into GenAI and agent workflows (see Technical Controls).
- Implementing robust testing and incident response (see Testing, Incident Response and Readiness for AI Trade Secrets Litigation).

Governance and Policy

Organizations should adopt an internal GenAI policy that:

- Prohibits the use of all GenAI tools other than secure, enterprise-grade tools that have been approved and procured by the organization.
- Requires personnel to use the GenAI tools only through accounts provided by the organization (not personal subscriptions).
- Expressly clarifies and reinforces that the organization's trade secret policy, information security and information technology policy and electronic communications policy (or equivalents) apply to the use of GenAI tools.
- Requires personnel to minimize the inclusion of trade secret information into GenAI tools unless strictly necessary and compliant with the trade secret policy.
- Instructs personnel to exercise data deletion functions within the GenAI tool to erase confidential or sensitive information from the GenAI tool's memory as soon as the information is no longer needed.
- Explains the legal issues around GenAI, the trade secrets' value, and the risk of losing trade secret protection.
- Instructs personnel not to store outputs that are potentially trade secrets in areas of the GenAI tool or the organization's systems that are accessible by other personnel who do not have a legitimate need to know the information or otherwise should not have access to it.
- Provides for recurring audits and refreshers to enforce employees' compliance obligations (see Training).

Companies should also identify their most sensitive trade secrets – often referred to as "crown jewels" – and document clear rules governing whether and how those assets may be used with AI tools. Not all information should be treated the same, and high-value assets may warrant stricter restrictions.

Organizations should also consider governance frameworks for AI agents specifically. As AI agents become increasingly capable of taking autonomous action across systems, policies should address:

- Limitations on the level of autonomy that the agents are enabled to exercise.
- Scope limitations on what data agents can access.
- Approval gates for agent-initiated actions involving sensitive information.
- Logging and audit requirements for agent activities.
- Periodic review of connector permissions and API access scopes.

For a sample policy, see Standard Document, Generative AI Use in the Workplace Policy.

GenAI Vendor and Model Provider Due Diligence and Contracts

Before procuring a GenAI tool, organizations should conduct targeted vendor diligence focused on how the provider actually handles sensitive information in practice. This includes verifying documented data retention practices, deletion workflows, data residency architecture, and whether customer inputs are excluded from model training or other secondary uses. Organizations should also assess how data is segregated across tenants and encrypted in transit and at rest.

Model-related diligence should examine how fine-tuned models and derivatives are stored and isolated, what telemetry or human review processes exist, and how access to customer data is logged and controlled. Finally, vendors' broader security posture should be evaluated through independent audit reports (such as SOC reports), vulnerability management processes, key management standards, and documented incident response procedures and notification timelines. The goal is not simply to rely on contractual language, but to understand and validate the vendor's operational safeguards before deployment.

Negotiating a carefully structured agreement – combined with technical validation of the vendor's controls – can materially reduce, although not eliminate, trade secret risk when using GenAI tools.

An enterprise-grade GenAI tool should be accompanied by an agreement that contains various protections at a minimum, such as:

- Robust confidentiality, privacy and cybersecurity restrictions and obligations that apply to the vendor with respect to the customer's inputs and outputs, including, where applicable, a regulatorily-compliant Data Processing Agreement or Business Associate Agreement.
- Express obligations not to use the customer's inputs or outputs for purposes of training any AI model or improving products and services.
- A right to require inputs and outputs to be permanently deleted within a certain timeframe.
- Clarification that the customer owns and will retain ownership of all right, title and interest in and to all inputs and outputs, together with language assigning the customer to any right, title or interest the vendor may have.

Where fine-tuning or model customization is involved, agreements should address ownership of fine-tuned weights, restrictions on model distillation, export controls on trained artifacts, and post-termination deletion or return of customized models.

In addition, GenAI vendors may be able to provide additional protections (potentially at additional cost), such as:

- "Zero data retention," meaning that no customer's inputs or outputs will be stored beyond the time at which the output is generated and delivered to the customer, although this would result in significant functionality limitations due to the GenAI being unable to retain context or history and therefore may not be a practical solution.
- A requirement that no human personnel of the vendor will review any of the customer's inputs or outputs (thereby avoiding the practical risk of individual personnel members viewing trade secret information and misappropriating it).
- A right to audit the vendor for compliance with its obligations and to require the vendor to formally attest or certify compliance.

For a checklist outlining key AI vendor due diligence considerations, see AI Tool Vendor Due Diligence Checklist.

Training and Personnel Management

Organizations should reinforce basic confidentiality discipline alongside AI governance. Proper training of personnel is essential, especially with respect to GenAI, given the broad spectrum of GenAI tools and their configuration options. Require mandatory training with attestations from all personnel that they have read and understand the GenAI policy. This has been demonstrated to increase compliance while also serving as an example of the "reasonable measures" taken by the organization to maintain the secrecy of its trade secrets.

Proper management also includes maintaining up-to-date non-disclosure agreements, enforcing strong on-boarding and off-boarding procedures, and promptly revoking access when personnel depart.

Technical Controls

Approved GenAI use should be governed through risk-based technical controls proportionate to the sensitivity of the data involved, the architecture of the deployment, and the operational role of the system within the enterprise. In higher-risk environments, organizations may route AI traffic through secure gateways that apply data-loss prevention controls, field-level blocking, or automated redaction to reduce inadvertent disclosure of sensitive information, for example.

Access to GenAI systems and related connectors should be tightly controlled. Strong identity safeguards – including single sign-on, multi-factor authentication, and role-based access controls – can be leveraged, with permissions narrowly scoped to defined business use cases. API credentials and connector tokens should be subject to least-privilege principles, secrets management, and periodic rotation. Where organizations deploy internally developed agents or workflows, code signing, version control, and controlled deployment pipelines can help preserve system integrity.

Retention and logging practices require particular calibration. Retention settings should be configurable and limited consistent with audit, compliance, and investigative needs. Logging should balance evidentiary and recordkeeping value against data minimization, with deliberate decisions about whether logs capture metadata, content, or both. Real-time monitoring and anomaly detection may be appropriate in higher-risk deployments to detect unusual data aggregation, connector misuse, or unauthorized access patterns.

Infrastructure-level safeguards remain foundational. These may include segmented development, testing, and production environments; encryption in transit and at rest with appropriate key management; artifact registries for internally developed models or agents; and integration with enterprise data-classification frameworks to ensure that access controls and data loss prevention policies are informed by the sensitivity of underlying information.

In certain highly sensitive contexts, vendor-side "zero data retention" configurations may be appropriate, such that inputs and outputs are not stored by the service provider after delivery to the user. At the same time, organizations should recognize the practical tradeoffs between, on the one hand, eliminating retention and, on the other hand, capitalizing on the functional benefits of GenAI retaining context and "memory" of prior inputs and outputs and preserving records.

Testing, Incident Response and Readiness for AI Trade Secrets Litigation

In the GenAI context, protecting trade secrets means not only preventing disclosure, but also being able to detect and respond swiftly when problems arise. Technical safeguards alone are not enough. Organizations should regularly test their GenAI systems and be prepared to respond quickly if sensitive information is exposed.

Red-team exercises can help identify risks such as prompt leakage or overbroad system access. Targeted security testing should confirm that permissions are properly limited and monitoring tools can detect unusual behavior.

Before anything goes wrong, organizations should define a clear preservation plan for AI artifacts and datasets, maintain immutable artifact registries and access logs, and document chain-of-custody procedures for models and training data. When a potential compromise is detected, the response should be immediate: revoke access, issue legal holds, and run an AI-specific triage to identify the affected models, data, pipelines, and users. Forensics should focus on model similarity testing, dataset fingerprinting, pipeline reconstruction, and cloud audit logs. Remediation may require key rotation, retraining on clean data, isolating tainted components, or – where necessary – proposing targeted technical carve-outs in negotiations or court. Companies should also keep ready an "injunctive relief package" that documents their reasonable secrecy measures, describes the protected assets, and frames damages based on avoided R&D and compute costs.

Global and Cross-Border Considerations

Trade secret protection is only as strong as the weakest jurisdiction. Organizations should harmonize secrecy classifications and AI controls across regions and assume that inconsistent practices in one market can undermine global protection. Where data localization

or transfer restrictions apply, AI systems should be deployed in regional instances using split-knowledge architectures to avoid centralizing sensitive datasets or parameters. Vendor terms and security annexes should be standardized globally to require secrecy-compliant deployments, consistent audit rights, and clear remediation and deletion timelines.

Recent regulatory developments underscore the need for cross-border vigilance. On January 1, 2026, several U.S. state AI laws took effect, including California AB 2013 (requiring disclosure of training data information) and the Texas Responsible AI Governance Act. A December 2025 executive order signals federal preemption efforts, creating ongoing regulatory uncertainty. Organizations should monitor these developments closely, as disclosure requirements may intersect with trade secret protection strategies. For more information on recent AI legal and regulatory developments in the US, see Practice Note, Developments in US AI Law and Regulation: 2026 Tracker.

Common Pitfalls – And How to Avoid Them

The most common failures in protecting trade secrets in GenAI environments are not exotic technical exploits, but ordinary governance gaps, including.

- **Shadow AI.** Well-meaning employees paste sensitive code, customer data, or financial information into unauthorized public LLMs or plug-ins. The remedy is straightforward: provide approved AI tools and prohibit use of all other AI tools, block unapproved AI tools to the extent technically possible, train users on prompt hygiene, and enforce gateway or proxy controls that block or redact high-risk inputs.
- **False sense of security.** Enterprise licenses and "no-training" representations do not eliminate internal misuse, overbroad connector permissions, or downstream disclosure. Vendor assurances are only one layer of protection; internal access controls and monitoring remain essential.
- **Leaky logs and retention creep.** Telemetry, analytics, and backup systems can quietly capture and preserve trade secrets far longer than intended. Redact sensitive fields where feasible, define clear retention schedules, and conduct periodic audits to ensure logging and backups remain aligned with risk tolerance and investigative needs.
- **Classification gaps.** AI deployments often outpace updates to data classification policies. Organizations should clearly define which categories of sensitive information may be used in approved AI systems, under what conditions, and in which environments.
- **Overbroad vendor rights.** "Improvements," "feedback," or "derivatives" clauses can inadvertently dilute ownership or exclusivity. Define these terms and the associated vendor rights narrowly, prohibit training on proprietary content where appropriate, and maintain clear ownership of fine-tuned models and derivative artifacts.
- **Uncontrolled evaluations.** Public benchmarks, conference submissions, and third-party testing can expose configuration details, datasets, or tuning approaches. Use confidential evaluation modes, obfuscate sensitive parameters, and disclose only what is necessary.
- **Blurring public and internal guidance.** Marketing materials, case studies, demos, and conference presentations that reuse real prompts or internal examples can unintentionally reveal trade secrets. Require pre-publication review and substitute synthetic or sanitized examples by default.
- **Agent sprawl.** As organizations deploy multiple AI agents with overlapping permissions, aggregate system access can exceed what any single employee would have. Conduct periodic access reviews to ensure agent permissions remain aligned with least-privilege principles and defined business purposes.
- **Uncontrolled fine-tuning or data reuse.** Proprietary datasets used to fine-tune models can embed sensitive information into model artifacts that are later shared or repurposed. Maintain strict governance over training datasets, model weights, and derivative outputs.