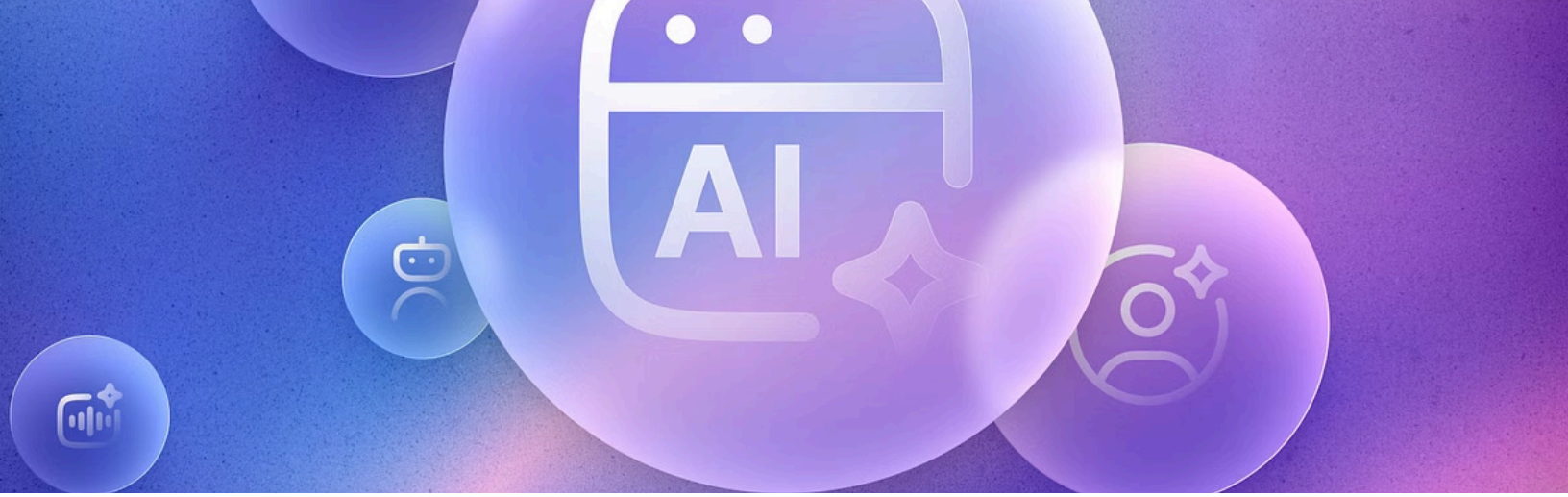

nudge

Practitioner's Guide to Agentic AI Security Four steps to get ahead of agentic AI risks

Agentic AI is already inside your organization.

Not in a proof-of-concept or a sandbox—in your SaaS tools, your employees' browsers, and your business workflows. And most of the time, nobody in IT or security approved it.

That's not a knock on your team. It's just how the Workforce Edge works. Over 90% of apps are now adopted by employees outside of IT—and agentic AI is following the same pattern, only faster. Employees are spinning up AI agents through tools they already use, connecting them to sensitive systems, and delegating their own access permissions in the process. Often with the best intentions, but almost never with a security review.

The result: A growing fleet of autonomous digital workers operating across your environment, each with persistent credentials and cross-system access, and most of them invisible to the people responsible for managing risk.

This guide breaks down what agentic AI means for your organization, why its autonomous nature fundamentally changes the security risk equation, and four concrete steps you can take right now to build governance practices that enable innovation while keeping your organization secure.

What agentic AI actually is

Agentic AI refers to AI systems—often built on or connected to large language models—that can take autonomous, multi-step actions toward a goal, not just respond to a single prompt. Instead of giving you an answer and stopping there, an agent can plan a sequence of steps to complete a task, choose which tools or applications to use, and execute actions on your behalf—sending emails, creating reports, updating databases, and more.

Think of it this way: If a large language model is the brain that predicts what to say next, an AI agent is the brain *plus* the hands and the job description. It has the tools, context, and instructions to act in the world.

This shift from ask-and-answer AI to goal-seeking autonomy is what has the security world paying attention. Autonomy changes the risk profile dramatically.

How agents connect to your systems

At their core, AI agents combine a large language model's reasoning capabilities with the ability to interact with external tools and systems. To do things—query a database, update a CRM record, pull information from internal documentation—agents need a way to connect to those business systems.

This is where the Model Context Protocol (MCP) comes into play. MCP servers act as bridges between the AI agent and your organization's applications, giving the agent structured access to the data sources and APIs it needs to understand context and execute tasks. An agent helping with customer support might use an MCP server to retrieve recent support tickets from Zendesk, pull account details from Salesforce, and reference internal knowledge base articles—all in real time, as part of a single workflow.

Without these connectors, agents are limited to their training data. With them, they become context-aware digital workers capable of autonomous, multi-step actions across your entire SaaS environment.

Behind every agent is a human decision

Agents don't spawn on their own. Behind every autonomous agent is an employee who signed up for a new AI tool, activated an agent feature in an existing SaaS platform, or connected an MCP server to a critical business system. That person configured the agent's permissions, pointed it to data sources like CRM records or internal wikis, and—often without realizing it—delegated their own access entitlements to it.

Most of the time, this happens with the best of intentions: save time, automate repetitive tasks, unlock a new business process. But it almost never comes with the oversight needed to assess the security or compliance implications of giving an autonomous system access to sensitive data, the ability to modify records, or the authority to act across multiple applications.

How agentic AI changes the risk equation

Traditional AI assistants wait for a user to act. Agentic AI flips the model: once triggered, agents can make independent decisions, sequence multi-step actions, and communicate with other applications or agents. That autonomy unlocks major productivity potential—but it also opens up pathways for attackers to manipulate agents into performing actions the human operator never intended.

A few factors make this especially tricky:

Agents don't work shifts.

They operate continuously, meaning a compromised or misconfigured agent can cause damage around the clock—not just during business hours.

Their behavior adapts.

Agents may change their approach based on conditions, making their actions harder to predict and audit.

They're often trusted too broadly.

Many organizations treat AI agents like internal applications and grant them high or persistent access privileges—without the same scrutiny they'd apply to a human employee with equivalent access.

This independence can lead to unintended, unauthorized, or risky outcomes—especially when governance guardrails are incomplete or absent.

Why your current governance approach probably won't cover this

Early AI governance frameworks were built for tools that assist—not tools that act independently across your SaaS environment. As agents move from proof-of-concept to production, most organizations are discovering that their existing governance models weren't designed for autonomous decision-making at this scale.

Here's where the blind spots tend to cluster:

Agent-driven shadow AI.

Just like shadow IT, employees are adopting agentic AI platforms outside formal procurement channels—introducing unmanaged third-party risk before security teams even know the tool exists. Nudge Security data shows the average organization already has 26 distinct AI apps in use. Most were adopted without IT oversight.

An expanded attack surface.

Agentic AI isn't just reading data—it's modifying systems, creating resources, and connecting to external services. Every OAuth token, API key, and cross-platform connection an agent holds is a potential entry point for attackers, especially if permissions are overly broad or poorly monitored.

Persistent access that outlasts its purpose.

Traditional AI assistants rarely touch your production systems. Agentic AI requires ongoing OAuth tokens or API keys—and if those aren't monitored and scoped properly, they become long-term liabilities. The average employee already creates 70 OAuth grants, 11 of which are considered high risk. Add agents to that picture and the problem compounds fast.

Multi-application workflows that cross compliance boundaries.

An agent might pull data from a regulated system and process it in another that isn't covered by the same compliance framework—creating inadvertent violations without anyone noticing.

Policies that haven't caught up.

From onboarding AI agents the way you'd onboard a new employee, to adding AI change management to your governance playbooks, enterprise policies need to evolve to address autonomous digital workers.

Fewer opportunities to intervene.

Autonomous execution reduces the windows in which a human can catch and halt a risky action before it's carried out. Real-time monitoring and automated guardrails matter more than ever.

Four steps to get ahead of agentic AI risks

The good news: you don't need to wait for a perfect governance framework to start protecting your organization. Many of the foundational security practices you already have in place—identity management, least privilege access, SaaS monitoring—apply directly to AI agents. You just need to extend them.

The key is recognizing that agents aren't just tools; they're digital workers operating with persistent permissions across your SaaS environment. That means applying the same rigor you'd use when onboarding a new employee: understanding what they have access to, what actions they're authorized to take, and how you'll monitor their activity over time.

1. **Get visibility into where agents are operating.**

   Build a real-time inventory of AI agents, their system connections, permissions, and who authorized them.

2. **Assess risks and misconfigurations continuously.**

   Evaluate access risk, data exposure, integration risk, and behavioral risk for each agent in your environment.

3. **Apply least privilege to programmatic access.**

   Audit OAuth grants and API keys, scope new agent access narrowly, and build a review cycle for agent credentials.

4. **Engage your workforce—don't just lock things down.**

   Make it easy to use AI agents safely, communicate clear policies, and treat workforce engagement as a security control.

Step 1

Get visibility into where agents are operating.

You can't govern what you can't see. That's true for shadow IT, and it's even more true for agentic AI—where the stakes of a blind spot are higher because the agent can take action, not just access data.

The first step is building a real-time inventory of AI agents operating in your environment. That means knowing:

- Which AI agents and agentic platforms are in use across your organization
- Which business systems they're connected to (CRM, code repositories, communication tools, financial platforms)
- What permissions they hold—OAuth scopes, API keys, MCP server connections
- Who authorized them, and under what circumstances

This is harder than it sounds. Agents often don't show up in traditional IT discovery processes because they're embedded in tools employees already use. A Salesforce user enabling an AI assistant in their account, a developer connecting an MCP server to their IDE, an operations manager activating an automated workflow in Notion or Slack—these all introduce agentic capabilities without triggering a procurement process.

What to do: Extend your SaaS discovery process to include agentic capabilities specifically. When you're reviewing OAuth grants and API connections—and the average employee has 70 of them, 11 of which are high risk—flag integrations that indicate agentic or automated access patterns. Prioritize connections to systems that hold sensitive data or have write access to production environments.

Step 2

Assess risks and misconfigurations continuously.

Inventory tells you what exists. Risk assessment tells you what to do about it.

The challenge with AI agents is that risk isn't static. An agent that's properly configured today can become a liability if its permissions drift, if the underlying model is updated, or if it's connected to a new data source it wasn't originally designed for. A one-time security review isn't enough.

For each agent in your environment, you want to understand:

Risk Type	What to assess
Access risk	Does the agent have more access than it needs? Agents frequently inherit the full permissions of the user who authorized them—meaning an employee with broad access to your CRM, code repositories, and communication tools can inadvertently grant an agent the same footprint. Over-permissioned agents are one of the most common and consequential misconfigurations you'll find.
Data exposure risk	What data can the agent access, process, or transmit? Agents connected to regulated data sources—health records, financial data, personally identifiable information—create compliance exposure the moment they move that data across application boundaries.
Integration risk	What other systems is the agent connected to, and have those connections been vetted? Every MCP server and API integration is another link in a chain that could be exploited. Attackers have already demonstrated the ability to manipulate agents through prompt injection—feeding malicious instructions through content the agent processes, causing it to take actions the human operator never intended.
Behavioral risk	Is the agent behaving as expected? Autonomous systems can produce unexpected outputs when they encounter edge cases or adversarial inputs. Without monitoring, you won't know when an agent has gone off-script until after something has gone wrong.

What to do: Build continuous monitoring for your AI agent inventory the same way you'd monitor for misconfigurations in your SaaS environment. Prioritize agents with write access to production systems, connections to regulated data, and broad OAuth scopes. Set up alerts for behavioral anomalies—agents making unusual numbers of API calls, accessing data outside their normal scope, or generating outputs that don't match expected patterns.

Step 3

Apply least privilege to programmatic access.

Least privilege is a foundational security principle, but it's consistently one of the hardest to enforce in practice—especially across the sprawling, decentralized SaaS environments most organizations operate today. Agentic AI makes this harder, and the consequences of getting it wrong are more severe.

Here's the core problem: when an employee grants an AI agent access to their accounts, they're often delegating their permissions wholesale. The agent doesn't get a carefully scoped, minimal set of capabilities—it gets whatever the employee has access to. That might include admin rights to your CRM, the ability to send email on behalf of the company, or read access to financial systems the agent has no business touching.

This is the agentic equivalent of handing a new contractor your master key because it was easier than figuring out which doors they actually needed.

Then there's the persistence problem. Unlike a human employee whose access is reviewed on a regular cycle, OAuth tokens and API keys used by agents often persist indefinitely. An agent set up for a project two years ago might still have active credentials to systems it no longer needs—and nobody's thought to revoke them.

What to do:

Audit existing programmatic access.

Pull a full inventory of OAuth grants and API keys across your environment and look for connections that are overly broad, connected to sensitive systems, or associated with AI tools. Prioritize any tokens with write access or admin-level permissions.

Scope new agent access narrowly from the start.

When employees or teams onboard new AI agents, push them toward minimum necessary permissions. Read-only access where possible. Specific scopes rather than broad admin grants. Time-limited credentials that require renewal rather than persistent tokens.

Build a review cycle for agent credentials.

OAuth tokens and API keys don't expire on their own. Put a process in place to review and rotate agent credentials on a regular cadence—quarterly at minimum for high-risk access, more frequently for agents connected to regulated data.

Watch for scope creep.

Agents that start with minimal permissions often accumulate more over time as users expand their capabilities. Monitor for permission changes and require re-approval when an agent's access footprint grows.

Step 4

Engage your workforce—don't just lock things down

If your first instinct is to block AI agents at the network level or issue a policy prohibiting their use, you're fighting the Workforce Edge instead of working with it. That approach has a well-documented track record: 69% of employees bypass security controls when they get in the way of getting work done (Gartner, 2022). AI adoption won't be any different.

The most effective governance programs don't try to stop employees from using AI agents. They make it easy to use them safely. That means employees know what's permitted, understand the risks of going outside those boundaries, and have a clear path to request access to tools they need. It also means security teams stop being the last line of defense and start being part of the workflow.

What to do:

Develop and communicate a clear AI acceptable use policy.

Employees who are experimenting with agentic tools often aren't trying to create risk—they're trying to get work done faster. Give them a policy that's specific and practical, not vague and restrictive. Cover which types of AI agents are permitted, what kinds of data they can and can't access, and how to request approval for tools that don't meet current criteria.

Meet employees at the moment of adoption.

The best time to guide an employee toward a compliant choice is when they're signing up for a new tool—not six months later when security discovers it in an audit. Browser-based guardrails and just-in-time prompts can surface your AI policy at the exact moment someone is onboarding a new agent, without requiring them to remember a training they attended once.

Make the approved path the easy path.

If getting approval for an AI tool takes two weeks and three forms, employees will find another way. Streamline your request and review process for AI agents so that working within the approved path is faster than going around it.

Build AI governance into your onboarding process.

New employees who understand your AI policies from day one are less likely to inadvertently introduce risk. Include agentic AI explicitly in security onboarding—not as a list of prohibited behaviors, but as a practical guide to using AI safely within your environment.

Treat workforce engagement as a security control.

An employee who understands why certain AI configurations are risky is more likely to flag concerns, ask questions before connecting a new integration, and make better decisions in edge cases your policy didn't anticipate. Security education isn't a soft add-on to your governance program—it's how you scale your controls across the entire organization.

A note on guardrails

As you work through these four steps, you'll start identifying places where manual processes aren't fast enough or consistent enough to keep up with the pace of AI adoption. That's where automated guardrails earn their keep.

Automated guardrails don't replace human judgment—they extend it. They can catch a risky OAuth grant before it's approved, surface a misconfigured agent before it touches sensitive data, and prompt an employee to review your AI policy at the moment they're considering a new tool. They can also help you enforce least privilege at scale, flagging overly broad permissions across thousands of agent credentials without requiring a manual review of each one.

The key is connecting your guardrails to the Workforce Edge—where the actual decisions about AI adoption are being made. Controls that only apply to managed devices or known applications won't catch the agents your employees are spinning up through browser-based tools and unmanaged SaaS accounts.

Getting started

Agentic AI governance doesn't have to be a moonshot project. Start with what you can see, extend least privilege to what you've found, and build employee engagement into your process from the beginning.

The organizations that get this right won't be the ones that locked AI down the tightest. They'll be the ones that figured out how to let their workforce move fast—and built the visibility, controls, and culture to keep that speed from becoming a liability.