

# The New Agentic Enterprise

## A CISO's Guide to Governing AI Agents without Stopping Them

Govern AI agents at runtime across browser sessions, applications, files, and data without slowing enterprise adoption.





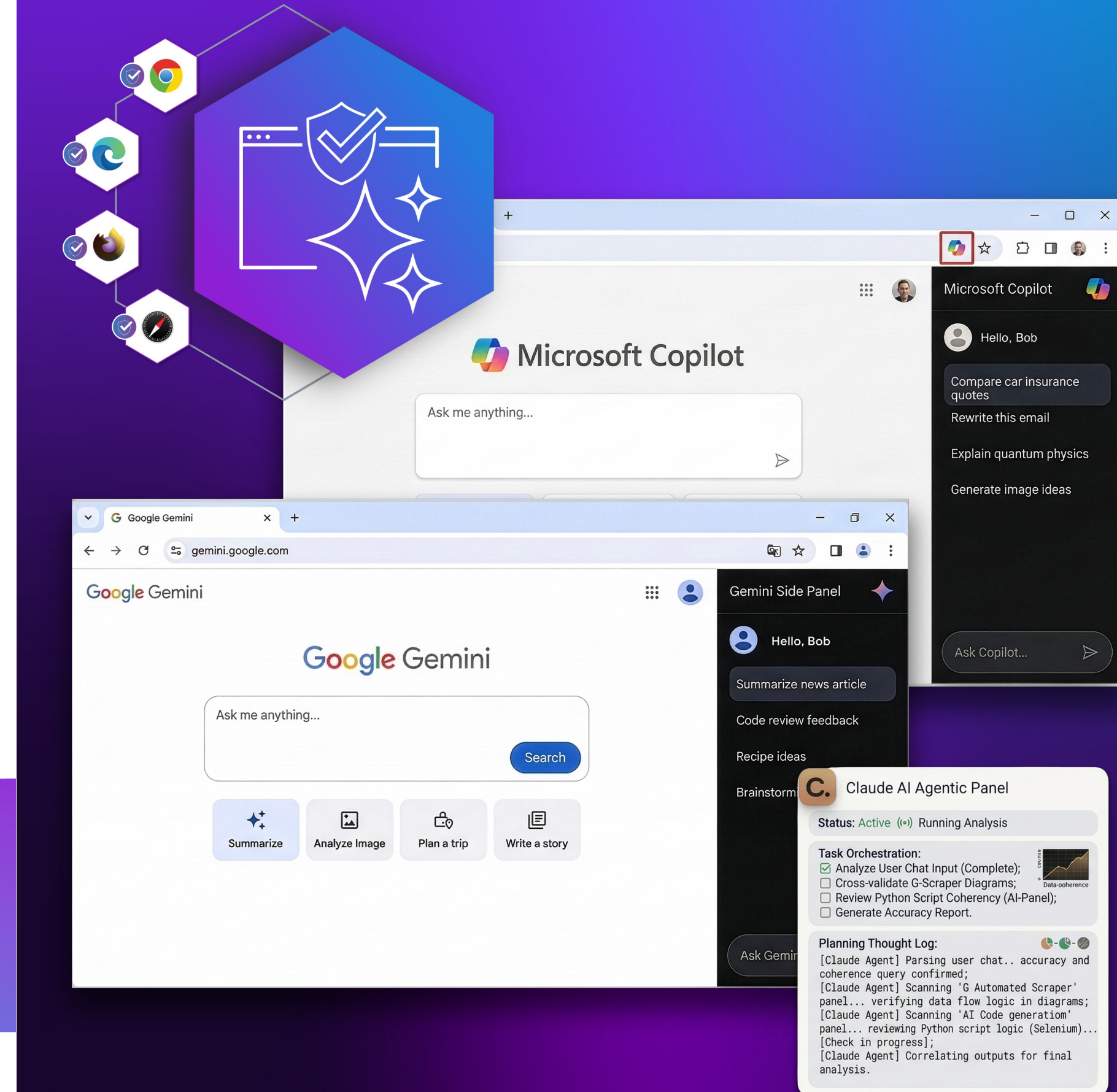
## Risk No Longer Sits Soley in the Model

AI agents are moving quickly from experimentation into everyday enterprise workflows. They can interpret objectives, access applications, move files, handle sensitive data, and carry out tasks with limited supervision. Risk no longer sits soley in the model. It emerges when an agent acts across the systems and data that the business relies on.

**That changes the security problem.** Existing identity, network, endpoint, and application controls remain essential, but they were designed for human-paced work. Agents can operate across multiple applications and repeat mistakes at machine speed, often before a flesh-and-bone user can intervene. Governance, therefore, has to persist beyond authentication and into the active session.

When it comes to AI agents, the browser is the practical control point. It brings together identity, application activity, prompts, files, data, and the sequence of actions. From there, enterprises can apply policy, detect prompt injection, inspect files, protect sensitive data, and introduce a human in the loop when the risk justifies it.

A workable governance model starts by identifying where agents operate, classifying the work they perform, defining permitted behavior, enforcing policy during execution, recording the context behind decisions, and reviewing access and ownership as agents evolve.



## Traditional Controls Were Built for Human Speed

AI agents are moving from experimentation into everyday business workflows. They can research, retrieve information, update systems, move files, and interact with enterprise data. Security leaders are being asked to support that adoption without creating risks that the organization cannot see or control.

Saying no to enterprise AI models may push employees toward unsanctioned tools, resulting in shadow AI.

**SAY NO**



**SAY YES,  
UNCHECKED**

Saying yes to AI without the right controls exposes agents as they begin acting across applications and data with limited supervision.

Unlike assistants that return a response, agents interpret goals and carry out tasks. A human may set the objective, but the agent may decide how to reach it, including entering data, uploading files, or triggering business processes. The risk appears when the agent acts. To govern that activity, security teams need visibility into the applications, data, files, and actions involved. That context comes together inside the browser session, where people and agents increasingly work.

**Enterprises must govern AI agents at the point where their decisions become business actions.**



## Agentic AI Has Changed How Enterprise Work Gets Done

Traditional workflows assume a person-in-the-loop approach, where real users remain involved at each meaningful step — users choose the application, review the information, decide what to do next, and confirm the result.

Agentic AI changes that sequence. An agent can interpret an objective, select tools, access applications, and carry out multiple actions with limited supervision. While the user defines the goal, the agent may determine the path—leaving it open to external risks such as prompt injection.

The same autonomy that speeds up a workflow can repeat a bad instruction across hundreds of records or systems before anyone notices. The speed that makes agents valuable can also magnify the impact of an unsafe file, inappropriate data transfer, or flawed instruction.

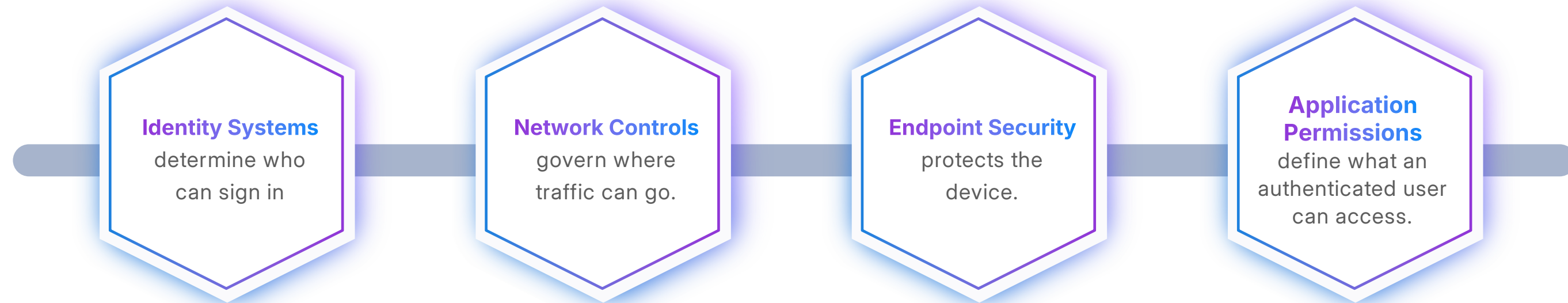


**AI risk cannot be treated as a mere model issue. It emerges across the full chain of execution: the identity behind the request, the applications the agent accesses, the files it moves, the data it handles, and the actions it delegates to other tools or systems.**



## Traditional Controls Were Built for Human Speed

Despite the proliferation of AI assistants, enterprise security controls still play an essential role in the enterprise tech stack.



These controls assume that a person remains present to interpret a warning, reconsider an action, or request approval. An AI agent may continue across multiple applications, handle files and data, and repeat the same action before a person can intervene.

## When AI Acts, Risk Becomes Real

Model security matters, but it cannot govern everything that happens after an AI system generates a response. Operational risk emerges when an agent turns that response into an enterprise action.

Governing the action requires more context than whether the agent was approved. It must answer who initiated the task, which agent is acting, what application and data are involved, whether the action is consistent with policy, and whether a human should review it before it proceeds.

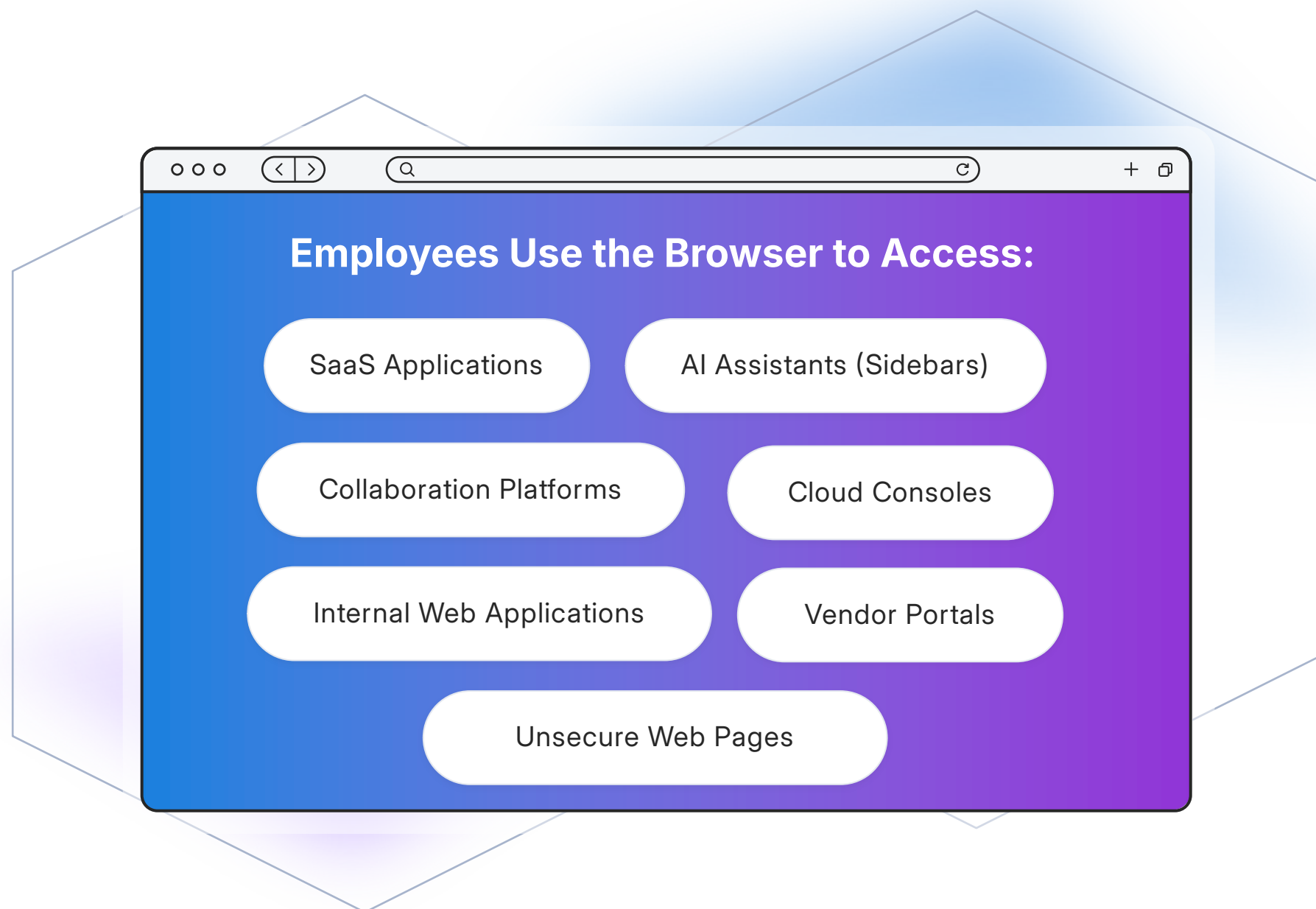
An approved agent can still take an inappropriate action. A legitimate request can still expose sensitive data, move an unsafe file, or trigger an unintended outcome. These risks span agent behavior, file security, and data security, and they can unfold at machine speed.

**For CISOs intent on avoiding non-compliance, this means governance must persist beyond deployment and authentication, with controls that inspect and mitigate risky actions during execution.**



## The Browser Is the Shared Execution Layer for People and Agents

By implementing enforcement at the browser point, security teams can govern people and agents without replacing the investments already in place.



The browser session brings identity, application activity, prompts, files, and data into one view. And AI agents are entering the same environment and acting across many of the same systems—with no guardrails or gutchecks to ensure a safe experience.

Network controls see the connection. Endpoint security sees the device. Identity systems confirm authentication. The session shows the work itself: the active application, prompt and form entries, copied data, file movement, and the sequence of actions.

Session context reveals whether an agent has encountered a prompt injection, moved an unsafe file, or attempted to expose sensitive data. These actions can unfold at machine speed, often faster than a person can recognize and stop them. Security teams, therefore, need the right tools to evaluate behavior in the moment, with enough context to distinguish routine activity from actions that require a human in the loop.



**Prompt injection** can hide malicious instructions in a webpage, application, or file, influencing an agent to take actions the user never intended — entering sensitive data, uploading or downloading a file, modifying a record, submitting a form, or calling another service.



## CISO INSIGHT 1:

## Govern Agents Like Users, Then Account for What Makes Them Different

AI agents operate with greater autonomy, speed, and scale than human users. A control model designed for people must therefore be extended to account for how agents interpret instructions, move between applications, handle files and data, and act before a person can intervene.



### Establish Identity and Ownership

Every agent should have a defined business purpose and a named owner. The organization should know which user, team, or process the agent represents, which applications it is approved to access, and how long it is expected to remain active.

That identity should persist throughout the workflow. If an agent downloads a file, transfers data, or triggers an application action, security teams should be able to trace the activity back to the agent and the person or process that initiated it.

**Without that chain of accountability, agent activity can quickly become difficult to explain, investigate, or govern.**



### Enforce Policy During the Session

Authentication confirms the identity used to enter an environment. It does not guarantee that every action taken afterward is appropriate. Policies should govern what the agent enters, receives, uploads, downloads, and transmits after authentication.

This is especially important when agents encounter untrusted web content or files that may contain hidden prompt injections. A legitimate agent operating through an approved identity can still be manipulated into exposing sensitive data or taking an action that conflicts with policy.

**These controls can inspect and mitigate risky activity before it becomes an enterprise event.**



### Apply Least Privilege to the Task

A user may have broad access because their role requires it across many workflows. An agent performing a single task may only need a small subset of those permissions. Its access should be limited by the application, operation, data set, task, and duration involved.

For example, an agent asked to summarize customer records may need read access to a specific data set. It should not also be able to export the database, change account permissions, or upload those records to an external service.

**Task-level access reduces the damage that can result from a bad instruction, a prompt injection, or an unintended action.**



### Use Proportional Human Oversight

Keeping a human in the loop does not mean requiring approval for every action. That would remove much of the value agents provide.

Instead, human review should be tied to risk. Routine, low-impact activity can proceed automatically. Additional scrutiny should be introduced when an agent attempts to access sensitive data, move a high-risk file, perform a destructive operation, behave outside its defined purpose, or take an action that conflicts with policy.

**Agent governance should preserve the familiar foundations of identity, access, policy, and accountability while accounting for greater autonomy and speed.**



## CISO INSIGHT 2:

# Focus Controls on the Moments Where Risk Concentrates

Agentic risk does not appear evenly across a workflow. It concentrates on the moments when an agent interacts with an application, enters or receives data, moves a file, or delegates work to another system.

Application access, data entry, file movement, and delegation are the points where an instruction becomes an operational action.



### Application Interaction

Security teams need to govern which applications an agent can access and what it can do inside them.

Access to an application should not automatically include permission to perform every available action. An agent may be allowed to view records without modifying them, submit information without changing settings, or retrieve data without exporting it elsewhere.

**Restricting the available operations limits the damage caused by prompt injection, a misinterpreted request, or another unintended action.**



### File and Data Movement

An uploaded or downloaded file may contain malicious content, embedded instructions, or hidden prompt injections intended to influence the agent. It may also carry sensitive information outside the organization through an approved application or workflow.

File security and data security, therefore, need to operate together. Controls should inspect files for unsafe content and evaluate whether the information they contain is appropriate for the destination.

**By extending file security into agent interactions, malware proliferation is prevented for both internal and external stakeholders, reducing overall enterprise risk.**

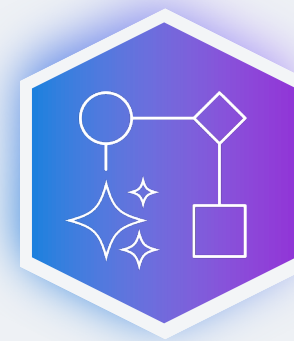


### Prompt and Data Entry

Prompts, forms, and text fields are common points of data exposure.

Before an agent enters information into an AI service or external destination, security controls should determine whether the content includes sensitive, regulated, or proprietary data. Policy should account for the destination, the identity behind the task, the agent involved, and the business purpose of the interaction.

**Where possible, sensitive data can be masked or restricted without blocking the entire workflow. This keeps productive activity moving while preventing information from reaching an inappropriate destination.**



### Delegation

Agentic workflows rarely remain within one application. An agent may pass a task to another agent, call a tool, retrieve data from one system, and submit the result to another.

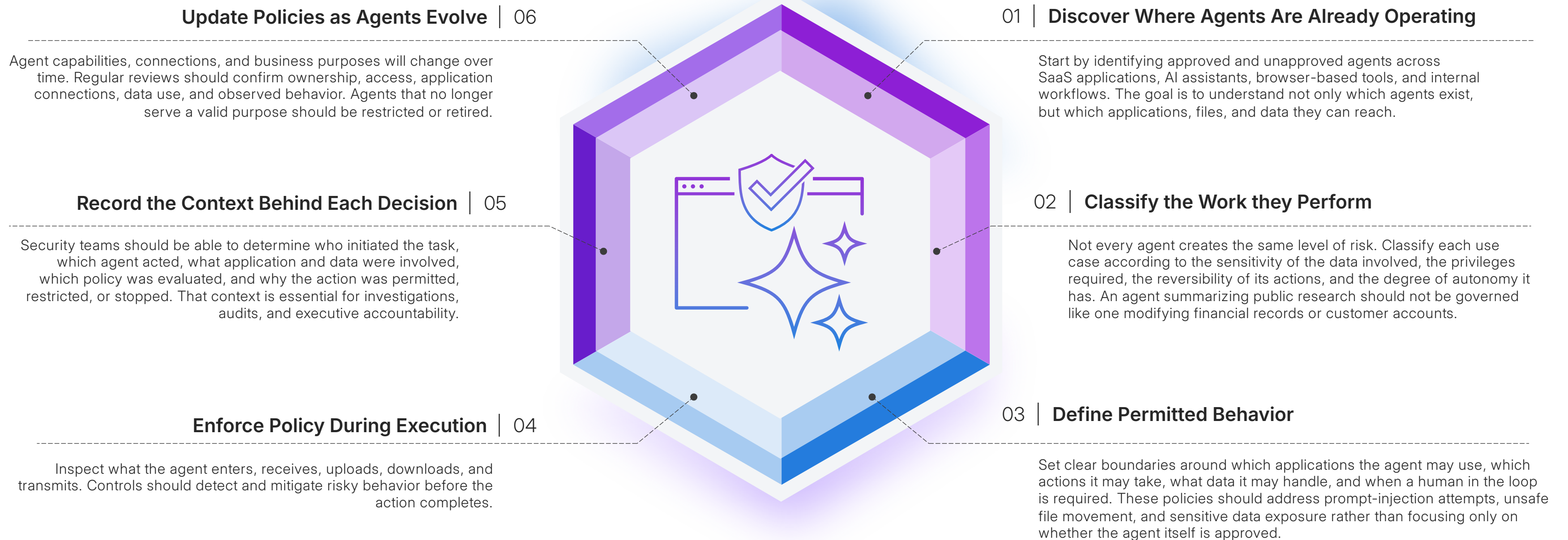
Each handoff can create a visibility gap if the organization cannot follow the chain of activity. Security teams need to preserve context as work moves between agents, tools, and applications, including who initiated the task, what data was transferred, and which actions were taken.

**CISOs must contend with recurring regulatory audits. Therefore, AI agent governance should also follow the full sequence of work, from the original instruction to the final outcome.**



## A Practical Governance Model for the Modern CISO

CISOs need a governance model that supports AI experimentation without allowing temporary pilots to become permanent sources of unmanaged risk. The model should be simple enough to apply across use cases, but specific enough to govern what agents can access, move, and change.





## Menlo Agent Runtime Security (MARS) Governs AI at the Point of Business

MARS is the governance layer for agentic AI, built on proven Menlo capabilities including the ability to isolate web traffic via the Menlo Cloud.

### Browser Security Platform

Agent-initiated web sessions run isolated in the cloud, so malicious content an agent touches never reaches the endpoint or the corporate network.



### Menlo File Security

Level 3 Content Disarm and Reconstruction, hash checking, antivirus, and sandboxing capabilities prevent zero-day threats in agent-accessed documents.



### Menlo AI Adaptive DLP

Sensitive data (PII, PCI, crown jewels, etc.) is automatically detected and masked in real time, ensuring that if hijacking occurs, data governance remains.



### HEAT Shield AI

Detects evasive web threats, including injection payloads embedded in pages and content crafted to manipulate agent behavior.



**MARS leverages these proven Menlo solutions to give CISOs a practical way to support agentic AI within the enterprise — without surrendering security, productivity, visibility, policy, or control.**



# See How Menlo Helps Organizations Govern and Secure the Use of AI Agents



| Your Agents Today |                                                                    | Your Agents on MARS |                                                                        |
|-------------------|--------------------------------------------------------------------|---------------------|------------------------------------------------------------------------|
| ✗                 | Agents inherit full user privileges, with no scoping to the task   | ✓                   | Per-agent identity, token authentication, and attribution              |
| ✗                 | Poisoned pages and files quietly redirect what the agent does      | ✓                   | Prompt injection and hidden content stripped before the agent reads it |
| ✗                 | Sensitive data leaves through the agent with nothing inspecting it | ✓                   | Adaptive DLP and masking on upload, download, and clipboard            |
| ✗                 | No record of what the agent actually did, or why                   | ✓                   | Session recording, per-agent dashboards, and human takeover            |

[Learn About Menlo AI Security](#)



Get in touch with us.  
[ask@menlosecurity.com](mailto:ask@menlosecurity.com)

© 2026 Menlo Security, All Rights Reserved.

