viLogics
Business-Class Cybersecurity. Risk, Fully Managed.

Securing AI Agents: 7 Controls for Agentic AI

Your AI agents have permissions. Who's controlling what they can do?

Artificial intelligence can improve productivity and strengthen security. Those same capabilities can accelerate fraud, expose sensitive information, and carry out harmful actions at scale.

Like nuclear energy, AI illustrates a fundamental truth: concentrated power creates extraordinary opportunity and demands serious safeguards. The outcome depends on how that power is governed and who controls it.

What the banking reports reveal

Recent reporting describes major U.S. banks blocking third-party AI agents from accessing client accounts and making transactions. Some agents reportedly attempted actions beyond user instructions.
Read the Reuters report.

These reports highlight delegated-access risks. They do not establish that AI has stolen money from millions of accounts. Giving an assistant credentials without reliable limits creates exposure.

Related news coverage

Watch the news segment on YouTube >

The risk extends beyond banking

Agentic AI pursues goals through connected tools and services. In email, payroll, client records, and cloud systems, its permissions become a security boundary. An agent authorized to review invoices should not automatically be able to change payment instructions.

What have we authorized AI to do, and what prevents it from doing more?

Set boundaries before agents act.

1. Approve agents before deployment

Inventory applications, agents, connectors, and tools. Assign an owner. Review permissions, data handling, and business purpose. Restrict unapproved deployments.

2. Limit access to the task

Use separate identities and narrowly scoped permissions. Start with read-only access. Avoid sharing banking passwords or broad administrative credentials.

3. Require independent approval

Keep people responsible for payments, new beneficiaries, payroll changes, sensitive exports, and destructive actions. An agent requesting an action should not approve it.

4. Protect data throughout the workflow

Apply policies to prompts, uploads, responses, and tool interactions. Protect credentials, financial information, client records, and confidential material.

5. Restrict destinations and tools

Define which services agents may contact and which tools they may invoke. Treat websites, documents, and connectors as potential sources of malicious instructions.

6. Monitor cumulative impact

Watch for unusual access, repeated small transactions, unexpected destinations, and permission changes. Connect activity to the agent and its responsible owner.

7. Be ready to stop an agent

Test how to disable agents, revoke sessions and tokens, preserve evidence, and contact affected providers before an incident occurs.

How viLogics turns AI security policies into tested controls

AI security policies only work when they can be enforced.

viLogics starts by understanding how your AI agents operate, what systems they access, what actions they're authorized to take, and where human approval is required.

We translate those business requirements into enforceable controls across supported security technologies, then test whether those controls actually work.

Our approach connects **business context, technical enforcement, and documented validation** so security teams have evidence that AI agents are operating within approved boundaries.

1. Map the permitted workflow

Identify the user, agent, business task, data sources, connected tools, and destinations. Define what successful authorized work looks like before deciding what to allow.

2. Write rules with business context

Example: Finance employees may use an approved assistant to summarize permitted invoices. Uploads to unapproved AI services are blocked when sensitive financial data is detected in inspected traffic. Exceptions require an owner and an expiration date.

3. Control agent connections

Inventory supported agent-to-tool communications, including Model Context Protocol (MCP). Allow approved tool servers and restrict unapproved connections. Restrict deployment through endpoint, identity, and application controls.

4. Inspect data and supported actions

Apply data loss prevention to inspected prompts, files, and tool exchanges. Where supported, block risky actions such as credential manipulation or destructive changes. Use runtime defenses to detect attempts to redirect an agent through malicious instructions.

5. Prove the rules work

Test approved work alongside attempted prohibited uploads, unapproved tool connections, and restricted actions. Verify the expected block, alert, and audit record. Tune policies when legitimate work is interrupted.

The deliverable: a workflow map, policy matrix, documented exceptions, validation results, and a response playbook. Coverage is agreed for the specific systems and integrations deployed.



Stop the action. Protect the data.

Example: an invoice assistant crosses the line

An approved assistant reads invoices. A document attempts to redirect it to export account details or change payment instructions. viLogics designs separate controls for each step rather than relying on the assistant to obey its original prompt.

Connection: Restrict calls to unapproved tools or destinations.
Data: Block detected sensitive information from leaving through inspected channels.
Action: Deny payment-write permissions for a read-only agent. Enforce independent approval for beneficiary changes in the financial application.
Response: Record the violation and trigger investigation; revoke access through the configured response process.

Pennies can change the economics (Hypothetical Scenario)

Imagine an attacker gaining access to execute unauthorized transactions across ten million accounts. At an average of just 50 cents per account, the cumulative loss would reach $5 million.

This is a hypothetical scenario, not a reported incident. Financial controls could prevent or detect such activity.

The larger concern is scale. Automated systems can repeat small, unauthorized actions thousands or millions of times, potentially turning minor individual losses into material financial exposure.

Security teams need to monitor cumulative behavior, enforce transaction limits, and detect patterns that individual transactions might not reveal.

The exposure goes beyond money

Client records, payroll files, credentials, and intellectual property also need protection. Data loss prevention applies across supported AI interactions, not only financial transactions.

Small, unauthorized actions can create significant losses before anyone notices.
Start the conversation with viLogics to map your AI workflows, identify risky permissions, and implement tested controls that protect your business.