vijilan
IT Security: Enabled

AI AGENT SECURITY GUIDE

Discovering and Securing AI Automation

Protecting the Next Generation of Automated Systems

AI-powered agents and automation tools are rapidly transforming how organizations operate. From customer support automation to workflow orchestration and intelligent data analysis, AI agents now interact with critical systems, applications, and sensitive data. While these technologies unlock efficiency and innovation, they also introduce new security risks. Unmonitored AI agents, insecure integrations, and uncontrolled access can create vulnerabilities that attackers may exploit.

This guide outlines how organizations can **discover, monitor, and secure AI-driven automation** while maintaining operational efficiency.

Understanding AI Agent Risks

AI agents operate differently from traditional software. They often interact with multiple systems, APIs, and data sources while making automated decisions. Without proper oversight, this can create new attack surfaces.

Common risks include:

Unmonitored Automation
AI agents executing actions across applications without proper logging or monitoring.

Excessive Permissions
Automation tools granted broad access to systems and sensitive data.

Prompt Injection & Manipulation
Attackers influencing AI models to execute unintended actions.

Data Leakage
Sensitive information exposed through AI responses or integrations.

Third-Party Integrations
AI tools connected to external services that may introduce supply chain risks.

Preparing for the Future of AI Security

AI adoption will continue to expand across every industry. Organizations that implement strong visibility, monitoring, and governance today will be better positioned to safely scale automation tomorrow.

Security teams must treat AI agents as active digital users within the environment—requiring monitoring, access control, and continuous oversight.

Discovering AI Agents in Your Environment

Many organizations adopt AI tools faster than security teams can track them. Visibility is the first step toward securing automation.

Key discovery strategies include:

- Identifying AI-enabled applications and platforms
- Monitoring API usage and automation workflows
- Tracking AI integrations with enterprise systems
- Identifying shadow AI tools used by employees
- Mapping AI agents interacting with sensitive data

Building a clear inventory helps security teams understand where AI agents operate and what systems they access.

Securing AI Automation

Organizations can reduce AI-related risks by implementing proactive security measures.

Monitor AI Activity
Continuously monitor logs, API activity, and system interactions generated by AI agents.

Implement Access Controls
Limit AI agent permissions to only the systems and data required.

Validate AI Inputs and Outputs
Detect prompt injection attempts, suspicious commands, or unexpected responses.

Secure Integrations
Review API connections and third-party integrations used by AI tools.

Maintain Audit Visibility
Ensure AI-driven activity is logged and available for investigation and compliance.

Security Best Practices

To strengthen AI security posture:

- Maintain visibility across all AI-driven automation
- Monitor system access and data usage by AI agents
- Detect abnormal behavior or unauthorized actions
- Implement strong authentication and least privilege access
- Continuously assess AI integrations and workflows