SecurIT360

AI SECURITY GUIDE · 2026

10 Things You Should Be Doing in AI Security

And why most organizations aren't doing any of them.

AI showed up in your organization whether you invited it or not. Your employees are using it, your vendors are embedding it, and attackers are weaponizing it. This guide covers 10 practical moves to use AI safely, and to defend against those who use it against you.

BY **Brad Causey,** VP of Offensive Security at SecurIT360
Daryl Irby, VP of Partnerships & Channels at SecurIT360

> AI risk isn't a project you finish. It's continuous, which means the organizations that manage it well are the ones watching it around the clock.

securit360.com · (205) 419-9066

SHADOW AI · AGENTS · AI ATTACKS · 24/7 SOC

WHAT'S INSIDE

The 10 Things, at a Glance

Most organizations are somewhere between "we blocked ChatGPT" and "we have no idea what our people are doing with AI." Neither is a strategy. AI is now part of your attack surface, your supply chain, and your workforce, and it's part of your adversaries' toolkit too. Here are the 10 things every organization should be doing right now, what most are doing instead, and how to close the gap.

THE 10 THINGS

What You Should Be Doing in AI Security

1 Find the Shadow AI: Inventory Your AI Usage

Before you can secure AI, you have to find it. Employees adopted AI tools long before most security teams had a plan for them: free chatbots, browser extensions, meeting transcribers, AI features quietly switched on inside SaaS apps you already own. Every one of them is processing your data. Most were never reviewed by anyone.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

You blocked one or two well-known AI sites at the firewall and called it handled. Meanwhile, surveys consistently find that a majority of employees use AI tools their employer never approved, and most of those users admit to putting sensitive data into them. Verizon's 2026 Data Breach Investigations Report found shadow-AI detections quadrupled in a single year.

Blocking one tool doesn't stop AI usage; it just moves it somewhere you can't see: personal accounts, personal devices, the next new tool that hasn't made your blocklist yet.

✔ WHAT YOU SHOULD BE DOING

Run a real discovery effort, then keep it current. Shadow-AI discovery isn't a one-time audit. New tools appear weekly, and existing vendors flip on AI features without asking. Your inventory should answer: what AI tools are in use, by whom, with what data, under what account?

- Review web proxy, DNS, and CASB logs for AI service traffic, including the long tail, not just the household names.
- Survey your business units. People will tell you what they use if the goal is enablement, not punishment.
- Check your existing SaaS stack for embedded AI features (CRM, meeting tools, office suites) that arrived via product update.
- Establish an intake path for new AI tools so the next request comes to you instead of around you.

2 Assign Ownership of AI Risk, Then Write the Policy

Everyone agrees AI risk matters. Ask who owns it, and the room goes quiet. Without a named owner, AI governance becomes everyone's side project and no one's job, and the policy, if one exists, becomes shelfware nobody enforces or updates.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Maybe there's a paragraph about AI buried in the acceptable-use policy from two years ago. Maybe legal drafted something nobody in IT has read. Nobody approves new tools, nobody tracks exceptions, and when an AI-related incident happens, the first hour is spent figuring out whose problem it is.

This is the gap we see most often: not a missing document, but missing accountability. The document is easy. Deciding who answers to leadership when AI goes wrong is the hard part, and it's the part most organizations skip.

✓ WHAT YOU SHOULD BE DOING

Name an owner first (a CISO, a vCISO, an IT director with a mandate), then build the policy under them. The policy itself should be short enough that people actually read it: which tools are approved, what data can go where, how to request new tools, and what happens when something goes wrong.

If you don't have a security leader who can own this, that's not a reason to skip it. It's the signal that you need one, even fractionally. This is exactly the kind of program-level gap a vCISO exists to fill.

- One named owner for AI risk, with authority to approve, deny, and revoke tools.
- A tool-approval workflow with an actual SLA. Slow approvals are how shadow AI wins.
- Policy reviews on a schedule (quarterly is realistic; the AI landscape moves fast).
- Leadership reporting: AI risk should appear in the same board conversations as ransomware and compliance.

3 Stop Sensitive Data From Leaking Into AI Tools

Every prompt is a data transfer. Source code, client records, financials, PHI, deal terms: once pasted into a consumer AI tool, that data has left your control, and depending on the tool's settings it may be retained, reviewed, or used to train someone else's model. This is the AI risk with the shortest fuse, and in regulated industries (legal, healthcare, finance), it's a confidentiality obligation, not just a best practice.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Employees use free-tier AI tools under personal accounts, where data-handling terms are weakest. Nobody has classified what's safe to share. The organization's "control" is a policy sentence nobody remembers, and surveys show the majority of unapproved-AI users admit to feeding sensitive data (customer details, employee records, internal documents) into these tools.

WHAT YOU SHOULD BE DOING

Give people a safe, approved path, then enforce the boundaries. Data leakage into AI tools is fundamentally a usability problem: if the sanctioned option is worse than the free one, people will use the free one.

- Provide enterprise-tier AI tools with contractual data protections: no training on your data, defined retention, admin visibility.
- Classify your data so "what can I paste?" has a clear answer. Green/yellow/red beats a 40-page standard.
- Extend DLP controls to AI destinations: browser-level controls, clipboard monitoring for high-risk groups, CASB policies for AI categories.
- Turn off training and history features where contracts allow, and verify the settings; don't assume defaults protect you.
- In regulated industries, map AI data flows against confidentiality obligations (privilege, HIPAA, GLBA) before an auditor or opposing counsel does it for you.

4 Vet Your AI Vendors

You're not just adopting AI tools; you're adopting the companies behind them. Where does your data go, who can see it, does it train their models, and what happens when they get breached? Most AI startups are optimized for growth, not security, and their data practices reflect it.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Tools got adopted because a free trial worked well. No security review, no contract review, no data-processing agreement. The vendor's privacy policy says they can use your inputs to "improve services," which means your data is now training data. And because AI vendors stack on top of each other (the app you bought is a wrapper on a model from someone else, hosted by a third party), your data may be flowing through subprocessors nobody evaluated.

WHAT YOU SHOULD BE DOING

Put AI vendors through the same third-party risk process as anyone else who touches your data, with a few AI-specific questions added. The goal isn't to slow adoption to a crawl; it's to know what you're agreeing to before your data is on someone else's servers.

- Does the vendor train on your data? Is opt-out contractual or just a settings toggle?
- Where is data stored and processed, for how long, and can you get it deleted?
- Who are the subprocessors? Which model providers and hosts actually see your prompts?
- SOC 2 / ISO 27001 attestations, breach-notification terms, and real security contacts.
- For embedded AI features in existing SaaS: re-review the vendor when the AI feature arrives. Your old assessment didn't cover it.

5 Lock Down AI Integrations

AI tools don't just receive what employees paste anymore. They connect. Copilots index your file shares, assistants read mailboxes, plugins and connectors reach into CRMs and ticketing systems through OAuth grants and API keys. An AI integration inherits every permission you give it, and it will faithfully surface everything those permissions can reach.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Broad OAuth scopes granted with one click. API keys that never rotate, scoped to everything. A Copilot rollout that suddenly made every over-shared SharePoint site searchable by every employee. The permission sprawl was always there; AI just gave it a search interface. Nobody reviews what's connected, and nobody de-provisions integrations when a pilot dies.

✓ WHAT YOU SHOULD BE DOING

Treat AI integrations like privileged accounts, because that's what they are. Least privilege applies double when the account holder can read and summarize a million documents in a minute.

- Audit existing OAuth grants and API keys tied to AI services; kill what's unused, narrow what's over-scoped.
- Fix permission hygiene before deploying enterprise copilots; they will expose every over-permissioned share you have.
- Scope AI service accounts to specific data sets, not tenant-wide access; rotate credentials on a schedule.
- Require security review for new AI connectors, the same as any other integration touching production data.

6 Govern Agentic AI Before It Governs You

Chatbots answer questions. Agents take actions: they browse, write code, send emails, execute workflows, and chain steps together without a human touching each one. That autonomy is the point, and it's also the risk: an agent with the wrong instructions, a poisoned input, or a compromised credential doesn't make one mistake. It makes mistakes at machine speed until something stops it.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Teams are piloting agents that touch production systems with no defined boundaries: what the agent may do, what it must ask permission for, what it can spend, what it can send outside the organization. Industry surveys through 2026 show the overwhelming majority of AI agents go live without full security approval; adoption is simply outpacing control. Prompt injection makes this concrete: an agent that reads external content (web pages, inbound email, documents) can be instructed by that content, meaning an attacker doesn't need your credentials if your agent will act on their words.

WHAT YOU SHOULD BE DOING

Give every agent the same treatment you'd give a new hire with system access: defined role, limited permissions, supervision, and an audit trail. The organizations getting this right constrain agents by design instead of trusting them by default.

- Human-in-the-loop checkpoints for consequential actions: payments, external communications, production changes, deletions.
- Scoped, short-lived credentials per agent, never shared service accounts with standing access.
- Full action logging: every step an agent takes should be reconstructable after the fact.
- Treat all external content an agent reads as untrusted input; assume prompt-injection attempts will happen.
- Kill switches: a documented, tested way to halt an agent immediately.

7 Don't Trust AI Output Blindly

AI output is confident by design, correct by coincidence. Hallucinated citations, fabricated facts, and insecure code all arrive with the same polished tone as the good stuff. The organizations getting burned aren't the ones using AI; they're the ones that removed human verification from the loop.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

AI-generated code ships to production with minimal review. Research testing 16 code models across 756,000 samples found that roughly one in five package dependencies suggested by AI coding tools are hallucinated: they don't exist. Attackers now register those predictable fake names in public registries and load them with malware, an attack known as "slopsquatting." Beyond code: AI-drafted client communications with invented facts, AI-summarized contracts missing key terms, fabricated case citations that have already sanctioned real attorneys in real courtrooms.

WHAT YOU SHOULD BE DOING

Keep a human accountable for everything AI produces that ships, sends, or decides. Verification isn't a rejection of AI; it's the control that makes AI safe to use at scale.

- Code review and SAST/SCA scanning apply to AI-generated code with no exceptions; verify packages exist and are legitimate before installing.
- Pin dependencies and use private registries or allowlists where feasible to blunt slopsquatting.
- Require source verification for AI-produced facts, figures, and citations before they reach clients, courts, or regulators.
- Make it policy: the human who ships AI output owns AI output.

8 Prepare for AI-Powered Attacks Against You

You aren't the only one with AI. Attackers use it to write flawless phishing at scale, clone voices from seconds of audio, generate deepfake video for wire fraud, and automate intrusions end to end. The FBI's 2025 Internet Crime Report logged more than 22,000 AI-related fraud complaints with losses exceeding $893 million, and reporting rates suggest the real number is far higher.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Your defenses still assume human-speed, human-quality attacks: phishing training that teaches people to spot typos that no longer exist, wire-transfer approvals that trust a familiar voice on the phone, and detection thresholds tuned for attackers who take days, even as CrowdStrike's 2026 Global Threat Report puts average eCrime breakout time at 29 minutes, with AI driving much of that acceleration.

A convincing deepfake of your CFO is no longer a sophisticated attack. It's a commodity service.

✓ WHAT YOU SHOULD BE DOING

Update your controls for AI-speed, AI-quality attacks, and accept that some will get through, which makes detection and response the control that matters most.

This is where around-the-clock monitoring stops being optional. When attackers move from initial access to lateral movement in minutes, a SOC that sees the alert tomorrow morning might as well not have seen it at all. If your team can't watch 24/7 (and most internal teams honestly can't), that's a capability to buy, not improvise.

- Out-of-band verification for wire transfers, payment changes, and credential resets. A voice on a phone is no longer identity.
- Refresh phishing and vishing training with AI-era examples: perfect grammar, cloned voices, deepfake video calls.
- Test your help desk against AI-assisted social engineering; it's a favorite initial-access path.
- Ensure 24/7 detection and response coverage: in-house if you can truly staff it, or through a managed SOC/MDR partner if you can't.

9 Log and Monitor AI Activity

Everything so far (inventory, policy, integrations, agents) depends on one capability: visibility. If you can't see what AI tools are doing in your environment, you can't enforce policy, catch misuse, or investigate incidents. And AI telemetry that nobody watches isn't monitoring; it's storage.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

AI tool logs exist somewhere, maybe, in each vendor's admin console, unreviewed. Agent actions aren't logged at all. When something goes wrong (data in a prompt that shouldn't have been, an agent that did something unexpected), there's no trail to reconstruct what happened. And no one is correlating AI activity with the rest of the security stack, so an attacker abusing an AI integration looks like normal traffic.

WHAT YOU SHOULD BE DOING

Get AI activity into your security-monitoring pipeline and put eyes on it. The goal is that AI usage, agent actions, and AI-integration traffic are visible in the same place your team watches everything else, with alerting for the events that matter.

Realistically, this is a staffing problem as much as a tooling problem. Collecting the logs is a project; watching them at 2 a.m. on a Saturday is an operation. This is precisely what a managed SOC does: ingest the telemetry, correlate it with endpoint and identity signals, and wake someone up when it matters. SecurIT360's SOC monitors client environments 24/7 (AI-related signals included), so the answer to "who's watching this?" is never "nobody."

- Centralize logs from enterprise AI tools, copilots, and agent frameworks into your SIEM.
- Alert on high-risk events: bulk data access by AI integrations, anomalous agent actions, AI-service logins from unusual locations.
- Retain AI activity logs long enough to support investigations; incidents are often discovered months later.
- Ensure someone is accountable for reviewing AI alerts around the clock, whether an internal SOC or a managed SOC, but named either way.

10 Put AI in Scope for Your Pentests and Assessments

If AI is in your production environment, it belongs in your testing scope. AI systems fail in ways traditional testing never looks for: prompt injection, jailbreaks, data extraction through crafted queries, agents manipulated into misusing their permissions. A pentest that walks past your AI systems is testing the environment you used to have.

⚠ WHAT MOST ORGANIZATIONS ARE DOING

Annual pentests still scope the same way they did five years ago: network, endpoints, maybe a web app. The new AI chatbot on the website, the copilot with tenant-wide data access, and the internal LLM app the dev team shipped last quarter have never been adversarially tested by anyone. Their first real test will come from an actual attacker.

WHAT YOU SHOULD BE DOING

Expand assessment scope to match your actual attack surface. Anything AI-powered that touches sensitive data or takes actions should be tested like the production system it is.

- Adversarial testing for customer- and employee-facing AI apps: prompt injection, jailbreaks, data extraction, abuse of connected tools.
- Include AI integrations and copilots in internal pentest scope: can a compromised low-privilege user leverage AI access to reach data they shouldn't?
- Add AI systems to your risk assessments and threat models, not just your tech-stack diagrams.
- Retest after significant AI changes: new models, new connectors, and new agent capabilities all change the attack surface.

BONUS

What Else Are You Possibly Missing?

Even if you're working through all ten, two things tend to get deferred, and both determine how well everything else holds up under pressure.

Update Your Incident Response Plan for AI

⚠ **WHAT MOST ORGANIZATIONS ARE DOING**

Your IR plan was written before AI. It has no playbook for "employee pasted the client database into a chatbot," no procedure for a rogue or compromised agent, no guidance for a deepfake-enabled fraud in progress. When one of those happens, your team will be improvising, on the clock.

✔ **WHAT YOU SHOULD BE DOING**

Add AI scenarios to your IR plan and exercise them: data exposure via AI tools, agent compromise, deepfake fraud, and vendor AI breaches. Define who gets called, what gets contained first, and what your notification obligations are when the "system" involved is a third-party model. Then tabletop it; an untested plan is a theory. Detection and response are two halves of the same capability: the same SOC that catches the incident at 2 a.m. should be feeding your response the moment it fires, which is why we pair 24/7 monitoring with digital forensics and incident response under one roof.

Train Your People on AI Risk

⚠ **WHAT MOST ORGANIZATIONS ARE DOING**

Your security-awareness program still teaches people to spot last decade's phishing. Nobody has told employees what's actually safe to put into AI tools, how to recognize a cloned voice, or what to do when an AI tool behaves strangely. The people using AI daily have received exactly zero minutes of training on it.

✔ **WHAT YOU SHOULD BE DOING**

Fold AI into security awareness now: what data can go into which tools (and why), how AI-era social engineering looks and sounds, verification habits for money and credentials, and a no-blame path for reporting AI mistakes. The goal is a workforce that uses AI confidently because they know where the edges are. Trained people are the cheapest control you'll ever deploy.

WHO IS SECURIT360?

Independent, vendor-agnostic security since 2009

Founded in 2009, SecurIT360 is an independent, vendor-agnostic cybersecurity and compliance consulting firm dedicated to safeguarding organizations' sensitive information. With offices in Birmingham, Alabama, and Kansas City, Missouri, our team of credentialed professionals delivers tailored solutions across a range of industries.

Who We Are

A team of cybersecurity experts committed to empowering organizations with the knowledge and tools to strengthen their security posture. Our independence keeps our recommendations objective and focused solely on our clients' best interests, and we tailor our services to each organization's specific goals and challenges.

What We Do

We help organizations identify, reduce, and manage cybersecurity risk through offensive security testing, strategic consulting, and 24/7 monitoring from our Security Operations Center. Services include managed detection and response, penetration testing, vulnerability assessments, compliance readiness (HIPAA, ISO 27001, NIST, and more), vCISO support, digital forensics and incident response, and security awareness training, all tailored to each client's environment to deliver insights that matter, not findings that check a box.

Our Mission

We are an organization focused on people. Our mission is to serve others by professionally and personally growing industry-leading experts who reduce the occurrence and impact of cyber crime. We do the right thing. We build relationships. We serve.

THE BOTTOM LINE

AI Security Isn't a Project. It's an Operation.

Here's the pattern across all ten items: policies, vendor reviews, and configuration are things you *set up*, but shadow-AI discovery, agent oversight, attack detection, and incident response are things you *operate*. They don't stay done. They require eyes on your environment every hour of every day, because AI has compressed attack timelines from weeks to minutes, and the incidents that matter will not schedule themselves for business hours.

Closing these gaps isn't a single purchase; it's a program. The good news: the ten things map onto three capabilities, and you don't have to build them all at once, or alone.

HOW SECURIT360 HELPS ACROSS ALL TEN

- **Govern & set up** (items 1–5). vCISO leadership, AI policy, third-party and vendor risk, and compliance readiness (HIPAA, ISO 27001, NIST) that put ownership and guardrails in place.
- **Validate** (items 7 & 10). Offensive security and continuous pen testing (PenGarde) that put your AI apps, integrations, and agents in scope before an attacker does.
- **Operate & respond** (items 6, 8 & 9). A 24/7 analyst-led SOC with integrated DFIR, watching AI-related signals alongside endpoint, identity, and network, then responding at machine speed.

Talk to our team about your AI security.

Wherever you land across these ten items, a conversation is a good place to start. We'll talk through what you already have in place, where the gaps are, and which capability — governance, testing, or monitoring — would move the needle most for your organization.

Talk to our team about AI security → securit360.com

or call (205) 419-9066