FREE GUIDE

The First 5 AI Security Fixes

Five plain-English actions to close the most common AI security gaps, in the order worth doing them.

Adding artificial intelligence (AI) to your products and workflows adds a few new ways for things to go wrong, but the first fixes are the same security discipline you already know, pointed at a new kind of system. You don't need a research background, and you don't have to do all five at once. Even the first one makes the rest easier.

A free guide from Sylvan Assurance — plain-English security, without us ever seeing your data.

Fix 1: Find the AI you already have (including the shadow AI)

What it is

A simple inventory of the artificial intelligence (AI) and machine-learning (ML) tools in use across your business: the ones you bought, the ones built into tools you already have, and the ones a team member signed up for without telling anyone ("shadow AI").

Why it's first

You can't protect what you don't know about. AI features arrive quietly: inside familiar tools, or through a free account someone opened to get a job done. Each one might touch customer data or act on your behalf. A list is the foundation every other fix builds on.

Do this

Write down every AI tool, assistant, and model in use, what data each one can see, and who owns it. Ask each team how they use AI day to day. You'll usually find one or two you didn't know about. About an hour, and it's free.

Fix 2: Control who and what can reach your AI systems

What it is

Putting the same access discipline around AI systems, model interfaces, and the keys that unlock them as you would around any sensitive account.

Why it matters

An AI system is often a doorway to data and actions. A leaked application programming interface (API) key, an over-shared model, or an assistant with more access than it needs turns a small mistake into a large one. Many AI incidents trace back to access that was broader than it needed to be.

Do this

Give each AI tool and key only the access it actually needs, store keys somewhere safe (not in code or chat), and rotate them on a schedule. Turn on two-step login for the accounts behind your AI tools.

Fix 3: Treat what goes into and out of a model as untrusted

What it is

Assuming that the text, files, and data a model reads (and the answers it gives back) could be wrong, manipulated, or unsafe, and handling them accordingly.

Why it matters

"Prompt injection" is a new version of an old problem: untrusted input changing what a system does. A cleverly worded document or web page can talk an AI assistant into leaking data or taking an action you didn't intend. And a model's answer can be confidently wrong. Neither should be trusted blindly.

Do this

Don't let a model take consequential actions (sending mail, moving money, changing records) without a human check. Keep untrusted content separate from your instructions, and don't paste secrets or customer data into tools that might train on them.

Fix 4: Vet your AI vendors and model supply chain

What it is

The same vendor due diligence you'd run on any supplier, applied to the third-party models, services, and AI features you build on.

Why it matters

When you use someone else's model or AI service, you inherit their security and their data practices. Where does your data go? Is it used for training? Who can see it? An AI vendor that's careless with your data is a risk you've taken on, whether you meant to or not.

Do this

Before adopting an AI vendor, ask where your data goes, whether it trains their models on your inputs, and how they handle a breach. Prefer vendors who let you opt out of training and who'll put their answers in writing.

Fix 5: Write a one-page AI use policy and turn on logging

What it is

A short, shared rule for how your team may and may not use AI, plus enough logging to see what your AI systems actually did.

Why it matters

Most AI trouble is a judgement gap, not a break-in: someone pastes customer data into a public chatbot, or trusts an answer they should've checked. A simple policy closes that gap. And if something does go wrong, logs are the difference between knowing what happened and guessing.

Do this

Write one page: what's fine, what's off-limits (for example, customer data in public tools), and who to ask when unsure. Turn on logging for your AI systems so you can review what they accessed and did. Read it over once or twice a year.

Where to go from here

The free AI Security Assessment shows where you stand across these areas (AI inventory, supply chain, prompt injection, data protection, access, monitoring, and governance) in about twelve minutes, and it runs entirely in your browser, so we never see your answers: sylvanassurance.com/free/ai-security/

When you're ready to turn that snapshot into a full programme with ready-to-use templates and runbooks, the AI Security toolkit lays it out: sylvanassurance.com/ai-security-assessment

Scan for the free assessment that goes with this guide.
sylvanassurance.com/go/free-ai-security-assessment