The First 5 AI Security Fixes

AI Security Assessment

Sylvan Assurance, LLC

Version 1.0 · Updated 2026-10-05

Contents

Fix 1 — Make the list: one page, every AI tool 1
Fix 2 — Read the training-data clause for your top three vendors 2
Fix 3 — Write the never-paste rule 3
Fix 4 — Cut the access list 4
Fix 5 — Decide your kill switch 4
Appendix — the five fixes on one page 5
Where you stand now 6

About This Guide **7**

A plain-English guide for the person who just became responsible for AI security.

Your organisation already uses AI. Not because anyone decided it would. It happened because a producer signed up for a transcription tool, a developer switched on a coding assistant, marketing started drafting copy in a chatbot, and the software you already pay for quietly grew AI features. None of it went through security review, because until recently there was nothing called "AI security review" to go through.

This guide is the starting point. Five fixes, in order, each one doable in a week or less of part-time attention. They are drawn from the most common gaps we see when organisations take our free AI Security Assessment, and from widely recognised AI security practice. None of them requires a budget, a consultant, or a data scientist. All of them are the kind of work you can show an insurer, a customer, or an auditor and say: we have started, and here is the evidence.

One honest note before we begin: these five fixes reduce common risks. They do not make you "secure"; nothing makes anyone secure as a finished state. What they do is move you from *unmapped* to *mapped*, which is where every defensible security practice starts.

Fix 1 — Make the list: one page, every AI tool

The gap. Ask yourself: could you produce, today, a written list of every AI tool in use across your organisation? Most people answer no, and the tools they forget are exactly

the ones that carry the risk: the free-tier accounts staff signed up for personally, the browser extensions, and the AI features that arrived inside software you already had.

Without an inventory, every other control is built on guesswork. You cannot read the terms of a tool you do not know about, revoke access you never recorded, or respond to an incident in a system nobody admits to using.

The fix. One spreadsheet, five columns:

Column	What goes in it
Tool	Name and vendor (include AI features inside other software)
Users	Who uses it: names or teams, including freelancers
Purpose	What it is used for, in one line
Data in	What categories of data go into it, and whether any is personal or confidential
Account type	Personal sign-up, free tier, or paid organisational account

Fill it in three passes:

1. **Survey.** Ask every member of staff (and every freelancer) what AI tools they use for work. Make it explicitly safe to answer honestly: the message is "we need the map", not "we are hunting offenders". People disclose when disclosure is welcomed; they hide tools when it is punished.
2. **Subscriptions.** Review card statements, expense claims, and your software subscription list for AI services.
3. **Features.** Walk your existing software stack (office suite, customer-relationship tools, design tools, meeting software) and note which have AI features switched on.

Done looks like: a dated, one-page register that you could hand to a new Information Technology (IT) lead on their first morning. It will be incomplete (every first inventory is) and that is fine. Incomplete and honest beats complete and imaginary.

Fix 2 — Read the training-data clause for your top three vendors

The gap. Somewhere in each AI vendor's terms of service is a clause that says what they may do with the data you put in, including whether your prompts, files, and outputs may be used to train or improve their models. It is the most consequential sentence in the contract, and it is the sentence almost nobody has read. These clauses also change over time, usually announced in an email no one opens.

If your client documents, customer records, or unreleased product plans are training someone else's model, that is not a hypothetical risk. It is a quiet, ongoing data transfer you have not assessed.

The fix. Take the three tools from your Fix 1 register that handle the most sensitive data. For each one, find and read three things:

1. **The training clause.** Does the vendor use your data to train, fine-tune, or "improve" their models? Is there an opt-out, and is it on by default for your account type? (Free tiers often have weaker protections than paid organisational tiers; sometimes the upgrade is the cheapest security control you will buy this year.)
2. **The retention terms.** How long are your prompts, uploads, and outputs kept? Can you request deletion?
3. **The sub-processor list.** Who else touches your data, including any third-party model provider behind the product?

Record the answers in your register with the date you checked. Where the answer is "yes, they train on our data, and there is no opt-out", you have a decision to make about that tool, and now you are making it with your eyes open.

Done looks like: three dated entries in the register recording what each vendor may do with your data, with any opt-outs exercised and any "we need to reconsider this tool" flags raised to whoever decides.

Fix 3 — Write the never-paste rule

The gap. The most common AI security incident is not an exotic attack. It is a person pasting something sensitive into a chat interface they should not have used: client contracts into a free summariser, patient details into a drafting tool, source code and credentials into a public assistant. A shadow-IT problem with an AI label on it, and it happens in organisations of every size, every week.

It happens because nobody ever wrote down the rule. Staff are not careless; they are unguided.

The fix. Write one page. Not a policy framework. One page, in plain language, sent to everyone and saved where people can find it. It needs four parts:

1. **What never goes into a public or unapproved AI tool.** Personal data (anything that identifies a living person), client or customer content, credentials and keys, and confidential business documents. List the categories that matter most in your organisation, with one concrete example each.
2. **Which tools are approved**, for what, and on which accounts (the organisational account, not personal sign-ups).
3. **How to ask for a new tool.** A named person or address, and a commitment to answer within two working days. If asking is slow or hostile, people stop asking.
4. **Where to report a worry**, including "I think I already pasted something I shouldn't have." Make it explicitly safe to self-report. The early report is the cheap version of the incident.

Send it to everyone, including freelancers and contractors. Keep a record of when it went out; the record is evidence of due care.

Done looks like: a dated, one-page acceptable-use note that every member of staff has received, and at least one new-tool request or self-report handled through the route it created. (The first self-report is a sign the rule is working, not failing.)

Fix 4 — Cut the access list

The gap. AI tool accounts accumulate access the way all accounts do, but with two aggravations: they are often outside the central login system (so the leavers process misses them), and what they can do is sometimes dramatic. A departed freelancer with surviving access to your voice-cloning tool, or a shared login whose password half the office knows, is an incident waiting for a motive. When it happens, you will not be able to say who did what, because everyone was the same user.

The fix. For every tool on your Fix 1 register:

1. **List who has access**: pull the user list from the tool itself, not from memory.
2. **Remove the leavers.** Anyone no longer with the organisation goes today. Then add AI tools to the standard leavers checklist so this never needs doing as a clean-up again.
3. **Kill shared logins.** One person, one account. If the tool's pricing pushes you toward sharing, treat that as part of the tool's real cost when you compare alternatives.
4. **Switch on multi-factor authentication (MFA)**, the second login step, usually a code from a phone app, for every AI account that supports it, starting with tools that hold sensitive data or can act on the world (send, publish, generate in your brand's voice).

Done looks like: every AI account attributable to one named, current person, MFA on wherever it exists, and AI tools written into your joiners-and-leavers checklist.

Fix 5 — Decide your kill switch

The gap. Some Thursday afternoon, an AI tool will do something bad: a model behind a product you rely on will start producing wrong or harmful output, a staff member will report that sensitive data went somewhere it should not have, or a tool will be misused by someone with access. The organisations that handle this well are not the ones with the thickest binders. They are the ones that decided, in advance, three small things.

The fix. Write one page, your AI incident starter card:

1. **Who is in charge.** One named person who makes the calls when an AI incident is suspected, and one deputy for when they are on holiday. Decision-by-committee is how the first hour gets lost.
2. **How to turn it off.** For each significant tool on your register: how do you suspend access, revoke the integration, or take the model out of the loop? For vendor tools this is usually an admin setting plus the vendor's support route; find both now, while nothing is wrong. Write down the steps. Containment for AI often looks like *roll back, disable the capability, or route around it* rather than "unplug the server".
3. **The first five steps.** (1) Stop the harm spreading: suspend the tool or access involved. (2) Preserve the evidence: export logs and capture screenshots before anything gets overwritten or deleted. (3) Write down what you know and when you knew it. (4) Assess whether personal data is involved; if it might be, data-protection notification clocks may apply, and this is the moment to confirm your obligations

with qualified legal counsel. (5) Tell the people affected what you know, plainly, when you know enough to be accurate.

Then spend one hour testing it: gather three people, pick a scenario ("our transcription vendor just emailed that customer audio was exposed"), and walk the card out loud. You will find two or three gaps. Finding them on a quiet Tuesday is the entire point.

Done looks like: a one-page incident card naming the owner, the off-switches, and the first five steps, and one tabletop hour on the calendar, done or booked.

Appendix — the five fixes on one page

Print this page and work down it. Dates in the right-hand column are your evidence of progress.

```
FIX 1 — THE LIST
[ ] Survey sent to all staff and freelancers              [date: _______]
[ ] Card statements and subscriptions reviewed            [date: _______]
[ ] Embedded AI features walked through                   [date: _______]
[ ] Register created: tool / users / purpose / data in /
    account type                                          [date: _______]
[ ] Register owner named: _______________                 [date: _______]

FIX 2 — THE TERMS
[ ] Top three tools by data sensitivity identified        [date: _______]
[ ] Training clause read and recorded for each            [date: _______]
[ ] Retention and deletion terms recorded for each        [date: _______]
[ ] Opt-outs exercised where available                    [date: _______]
[ ] "Reconsider this tool" flags raised where needed      [date: _______]

FIX 3 — THE RULE
[ ] One-page acceptable-use note written                  [date: _______]
[ ] Never-paste categories listed with examples           [date: _______]
[ ] Approved tools + request route + report route named   [date: _______]
[ ] Sent to everyone incl. freelancers; record kept       [date: _______]

FIX 4 — THE ACCESS
[ ] User list pulled from each tool                       [date: _______]
[ ] Leavers removed                                       [date: _______]
[ ] Shared logins eliminated or scheduled out             [date: _______]
[ ] Multi-factor authentication on where offered          [date: _______]
[ ] AI tools added to the joiners-and-leavers checklist   [date: _______]

FIX 5 — THE KILL SWITCH
[ ] Incident owner + deputy named                         [date: _______]
[ ] Off-switch steps written per significant tool         [date: _______]
[ ] First-five-steps card written                         [date: _______]
[ ] One-hour tabletop run (or booked)                     [date: _______]
```

When all five blocks are dated, re-take the free assessment. Your score will have moved, and more usefully, the priority actions it recommends next will have changed.

Where you stand now

If you have done these five fixes, you have: a map of your AI estate, eyes-open knowledge of what your key vendors do with your data, a rule everyone has seen, an access list with no ghosts on it, and a first hour you could actually execute. That is genuinely ahead of most organisations, and it is the foundation that everything more sophisticated builds on.

To see your full posture: the free AI Security Assessment scores you across eight areas (inventory, vendor and model supply chain, prompt injection and input handling, data protection, access control and agentic systems, monitoring and drift, incident readiness, and governance) in about twelve minutes, entirely in your browser. Your answers never leave your device.

To go further: the paid editions of the AI Security Assessment include the working documents this guide points toward, the full inventory register, the acceptable-use policy template, the vendor due-diligence questionnaire, the AI incident runbooks, and, at the top tier, the ten-dimension maturity rubric and board reporting pack. The companion Sylvan Press books (the *AI Security Field Guide* and the *AI Security Operator's Workbook*) carry the full reasoning and the complete template library.

Provided by Sylvan Assurance, LLC. This is general, plain-English guidance, not professional advice, not a security audit, and not legal advice. Every recommendation is optional; verify anything here against your own circumstances and, where needed, a qualified professional. Responsibility for your organisation's security remains with you.

About This Guide

This guide is produced by **Sylvan Assurance, LLC** and reflects professional experience in the relevant practice area, translated into plain-language guidance for small business owners.

It provides **general guidance and recommended practices**. It is **not** legal advice (and not a professional audit). Every recommendation is **optional** - a widely accepted way to reduce a common risk - and following it reduces exposure to common problems but **does not guarantee the outcome the product claims**. Responsibility for your business remains with you.

For full terms, see the Terms of Use provided with the AI Security Assessment.

Copyright 2026 Sylvan Assurance, LLC. Not for resale.

Free to share unchanged with credit. Terms: sylvanassurance.com/licence/free/

Where to go from here

Scan for the full toolkit that builds on this guide.
sylvanassurance.com/go/toolkit-ai-security-assessment

Scan for the companion book - the full story, in plain English.
sylvanassurance.com/go/ai-security