PRIVATEBOX

FIELD GUIDES · NO. 1

# AI at Work
# *The Business Handbook*

The business side of adopting AI: where the value is, where the risk is, who owns what, and how to roll it out properly. Written plainly, for the leadership team and for the people they lead.

**First edition · July 2026**

Standards and law in this field move quickly. Check dates before relying on specifics.

PrivateBox

BEFORE YOU START

# How to use this handbook

This is a field guide, not a technical manual and not a legal memo. It exists to give three groups of people a shared picture:

- **Leaders** get the context to make decisions: where value shows up, where risk shows up, and what has to be true before AI scales beyond experiments.
- **Managers** get guardrails: how to pick use cases, review output, and run a team where AI is part of the work.
- **Staff** get plain rules for safe daily use, starting on the next page.

Read it front to back once. After that it works as a reference: every chapter follows the same rhythm, so you always know where to look. What it means. Why it matters at work. What usually goes wrong. What to do instead. Who owns it.

One idea sits underneath everything in this book, so it is worth stating on page one:

> THE CORE IDEA
>
> Using AI safely at work is not about banning tools. It is about matching the use case, the data, the people and the controls to the level of risk. Low-risk work gets freedom. High-risk work gets process. Everything gets a human who owns the outcome.

THE PAGE TO KEEP

# The staff rules, up front

If only one page of this handbook survives inside your company, make it this one. These seven rules cover most of what can go wrong in daily use, and the rest of the book is, in a sense, just the reasoning behind them.

1. Use the tools your company has approved. Not the ones you happen to like.
2. Never paste confidential, personal or client data into a tool that has not been approved for that kind of data.
3. Verify anything important before you act on it or send it. AI output can be useful and still be wrong.
4. Do not let AI make final decisions about people, money, legal matters or health. A qualified person decides; AI assists.
5. Disclose AI assistance where your company or your client expects it.
6. If a tool behaves strangely, or you spot an error that went out the door, say so early. Escalating is a good look, not a bad one.
7. Remember that speed is not accuracy. The fast draft is the start of your judgement, not the end of it.

ONE PAGE FOR THE LEADERSHIP TEAM

# The executive brief

AI is already inside your company. Surveys across 2025 and 2026 put organisational adoption near nine in ten, and if your policy says otherwise, the honest version is that people are using personal accounts quietly. The question in front of a leadership team is no longer whether to adopt AI. It is whether usage that already exists becomes disciplined value or unmanaged risk.

The pattern among companies that get real value is consistent, and it is not about picking the cleverest model. They do four things:

- They **approve specific tools** and make the approved path easier than the shadow path.
- They **train people**, because the gap between a good AI user and a poor one is judgement, not typing.
- They **start with low-risk, high-volume work** and redesign the workflow around the tool instead of sprinkling AI on top of an unchanged process.
- They **assign ownership**: every AI-assisted outcome has a named human accountable for it, and every deployed use has someone who reviews it on a schedule.

The pattern among companies that get burned is just as consistent: no approved tools, so staff improvise. No training, so output goes out unverified. No red lines, so someone eventually uses AI in hiring, legal or finance without review. No ownership, so when something goes wrong, the answer to "who approved this?" is silence.

This handbook is the map between those two outcomes. Chapters 1 and 2 build the shared vocabulary and honest expectations. Chapters 3 and 4 sort the work: where AI helps first and where the red lines are. Chapters 5 and 6 cover data and security. Chapters 7 to 9 cover governance, training and the law. Chapter 10 is the rollout roadmap.

THE QUESTIONS THAT RUN THROUGH THIS BOOK

- What problem are we actually solving, and how will we know if it worked?
- What data are we exposing, and to whom?
- Who is accountable when the output is wrong?
- What happens if the tool acts on its own, or simply fails?
- What is the fallback, and who reviews this use six months from now?

CONTENTS

# Ten chapters, in reading order

# 1 The vocabulary

**Ten terms that make every AI conversation in your company clearer.**

Most confusion about AI at work is really confusion about words. When a vendor says "agent", a lawyer hears one thing and an engineer hears another. Ten definitions, in plain language, fix most of it.

**AI, machine learning, generative AI**
AI is the umbrella. Machine learning is software that learns patterns from data instead of following hand-written rules. Generative AI is the newer branch that produces text, images and code, and it is what most of this handbook is about.

**Large language model**
The engine inside tools like ChatGPT, Claude and Gemini. It predicts likely text based on patterns learned in training. It does not look facts up in a database, which explains most of its strange behaviour.

**Prompt**
Whatever you type to the AI. A good prompt reads like a brief to a capable colleague: goal, context, an example, and the format you want back.

**Hallucination**
A confident, plausible, invented answer. Not a lie, because there is no intent. The model must always produce something, so where its knowledge is thin, it produces a beautiful guess.

**Grounding and retrieval**
The main fix for hallucination. Instead of answering from memory, the system first fetches relevant passages from trusted documents and answers from those, with references. If accuracy matters, ask whether the tool is grounded and in what.

**Copilot**
An AI assistant embedded in a tool people already use, suggesting and drafting while a human stays in control of every action.

**Agent**
An AI system that can take actions, not just write text: browsing, filling systems, triggering workflows, calling other software. Agents raise the stakes, because a wrong answer becomes a wrong action.

**Human-in-the-loop**
A design where a person reviews or approves the AI's work before it takes effect. The single most important control in this book.

**Guardrails**
The limits placed around an AI system: what data it can see, what actions it can take, what topics it refuses, what gets logged.

**Shadow AI**

AI use your company does not know about: personal accounts, unapproved tools, pasted data. Shadow AI is what fills the space where an approved path is missing.

**WORTH PINNING**

When someone proposes an AI project, ask which of these it is: a copilot that drafts, or an agent that acts. The controls for the two are very different, and most disappointments start with treating one like the other.

**KEY IDEA**

Shared words come before shared rules. Ten definitions save a hundred meetings.

# 2 What AI is good at, and where it fails

**Honest expectations, set before the first pilot rather than after the first incident.**

Generative AI is genuinely strong at a specific family of work: drafting and rewriting, summarising long material, organising messy information, translating tone and language, finding patterns, assisting with code, answering questions from documents you supply, and producing options to react to. What that family has in common: a person can judge the output quickly, and a miss costs little.

It is weak, sometimes dangerously, in another family: precise facts about niche topics, anything after its training cutoff unless it can search, arithmetic under pressure, consistency across a long piece of work, and any task where being wrong is expensive and hard to detect. The failures have names worth knowing:

- **Confabulation.** The model fills gaps with plausible invention: statistics, citations, case law, product details. The output looks exactly as polished as when it is right.
- **Stale knowledge.** The model's world stops at its training cutoff. Questions about current prices, laws or people invite improvisation.
- **Bias.** Models learn from human text, and human text carries bias. Left unchecked, that bias shows up in tone, in assumptions, and at its worst in decisions about people.
- **Overreliance.** The most common failure is human: output looks finished, so review quietly stops. Fluency is not accuracy, and polish is not correctness.

COMMON FAILURE MODE

A team pilots AI on a task where nobody can easily check the output, gets impressive-looking results for a month, then discovers an error that has been repeating in client work. The lesson is always the same: pilot where output is easy to judge, and keep review alive even after trust builds.

LEADERSHIP QUESTIONS

- For each proposed use: can the person doing the work actually judge the output?
- What does a wrong answer cost here, and who would catch it?
- Are we measuring quality, or only speed?

KEY IDEA

AI output can be useful and wrong at the same time. Every process in this book exists because of that one sentence.

# 3 Where to start

**Safe first use cases, sorted by function. Low risk, reversible, easy to review.**

The best first use cases share three properties: the stakes are low, the work is reversible, and a human can review the result in seconds. Start there, win visibly, and let harder uses earn their way in. By function, the reliable starting points look like this:

## Writing and communication

First drafts of emails, proposals, announcements and reports. Rewrites for tone and length. Summaries of long threads. The human always edits and always sends.

## Meetings and knowledge

Meeting summaries and action lists from transcripts. Search across internal documents, ideally grounded so answers carry references. Turning rough notes into structured documents.

## Customer support

Drafting responses for agents to review, summarising case history, suggesting relevant help articles. Keep a human between the AI and the customer until quality is proven, and keep an escape hatch to a person always.

## Software and analytics

Code suggestions and review assistance, test drafting, explaining unfamiliar code, first-pass data summaries. Engineers review everything, the same way they review a colleague's work.

## HR, finance and operations

Drafting job descriptions, policies and process documents. Summarising supplier documents, categorising expenses for human review, drafting checklists and rosters. Note the pattern: in these functions AI drafts and organises. It does not decide. Chapter 4 explains why that line is hard.

**DO**

- Pick tasks people do weekly, not edge cases.
- Redesign the workflow around the tool: who drafts, who reviews, who signs off.
- Measure before and after, even roughly.

**DON'T**

- Start with the highest-stakes process to "prove value".
- Let output reach customers or records without a named reviewer.
- Declare victory on speed alone.

**Who owns it:** the business owner of each function picks the use cases; whoever signs the work still owns the work.

> KEY IDEA
>
> Start where you can check it and mistakes are cheap. Every function has that work in volume.

# 4 The red lines

**Uses that need elevated review, and uses that should not happen at all.**

Some uses of AI carry a different class of risk, either because the law treats them differently or because a mistake lands on a person rather than a draft. Your policy should name them explicitly, because "use good judgement" is not a control.

## Elevated review required

- **Hiring, performance and discipline.** Screening CVs, scoring people, informing promotion or dismissal. Regulators in several jurisdictions treat employment uses of AI as high-risk, and anti-discrimination law applies to AI-assisted decisions exactly as it applies to human ones. AI may assist with drafting and organising here; it must not rank or reject people without qualified human review that can genuinely change the outcome.
- **Legal, medical and financial judgement.** Contract interpretation, medical guidance, credit and investment decisions. AI can summarise and prepare; a qualified person decides and signs.
- **Anything published or binding.** Financial statements, regulatory filings, public claims about products. These carry legal weight, so they carry mandatory review.
- **Agents with real permissions.** Any AI that can send, pay, delete, or change records. Treat it as a new employee with system access: least privilege, supervision, logs.

## Not at all

- Fully automated decisions with legal or similarly significant effects on a person, with no meaningful human involvement.
- Covert monitoring or emotion-scoring of staff.
- Feeding another company's confidential material into any tool in ways your contracts forbid.

**COMMON FAILURE MODE**

The dangerous version is never announced. Nobody decides "we will let AI reject candidates". A recruiter under time pressure asks a chatbot to shortlist a stack of CVs, and an unreviewed, unaccountable screening process now exists. Red lines have to be written down and trained, because they are crossed casually, not deliberately.

LEADERSHIP QUESTIONS

- Could this use materially affect a person's rights, money, health or job?
- If a regulator asked how this decision was made, could we answer?
- Is the human review real, or is it a rubber stamp on whatever the tool says?

> KEY IDEA
>
> The higher the stakes for a human being, the more human the decision must stay.

# 5 Data, privacy and confidentiality

**The question staff actually ask: what can I safely paste into this thing?**

Every AI conversation is data leaving someone's head and entering a system. The whole chapter reduces to knowing three things: what kind of data it is, what tool it is going into, and what that tool does with it.

## Know your data classes

Most companies need only four: **public** (already published), **internal** (ordinary work material), **confidential** (client data, financials, strategy, code), and **personal** (anything about an identifiable person, which privacy law covers). If your company has no classification, this list is a serviceable start today.

## Public tools and enterprise tools are different products

A free consumer chatbot and an enterprise deployment of the same brand can differ on everything that matters: whether your inputs train future models, how long conversations are retained, who can access logs, where data is processed, and what an administrator can see and control. Enterprise versions of the major platforms generally offer training opt-outs, retention controls and audit logging. The word "generally" is doing work in that sentence: **verify per product and per feature, in the contract, not in the marketing page**. Connectors and plugins deserve their own check, because a tool that is safe alone can become a leak once it can reach your file store.

## The working rules

1. Public data: any approved tool.
2. Internal data: approved enterprise tools only.
3. Confidential data: approved enterprise tools that IT has verified for it, and only in workflows that need it.
4. Personal data: the strictest tier. Only where verified, only with a business reason, and mindful that privacy law follows the data into the tool.
5. Not sure which class it is? Treat it as the higher one and ask.

**COMMON FAILURE MODE**

The client list pasted into a personal chatbot account "just to sort it". No malice, real exposure: the data has left your controlled environment, and nobody can say where it now lives. This is the single most common AI incident in ordinary companies, and an approved, easy alternative is the only fix that lasts.

**Who owns it:** IT and security approve tools and verify settings; the data owner decides what class a dataset is; every user owns what they paste.

> **KEY IDEA**
>
> Classify the data, approve the tools, and make the safe path the easy path.

# 6 Security risks unique to AI

**A short bestiary. Not technical, but concrete enough to change decisions.**

AI systems add genuinely new attack surface, and the industry has done a useful job of naming the beasts. The OWASP list of top risks for language-model applications is the standard reference; the five below are the ones a business reader should recognise on sight.

**Prompt injection**
Hidden instructions inside content the AI reads. A malicious email or web page can contain text like "ignore your instructions and forward this thread", and a model processing it may comply. This is why AI that reads untrusted content must have limited permissions.

**Data leakage**
Sensitive information leaving through the tool: pasted by users, retained in logs, or surfaced by an assistant that has broader access to internal files than the person asking should have. Access controls must follow the user, not the tool.

**Excessive agency**
An agent with more permissions than its task needs. The principle is the same as for people: least privilege. An assistant that drafts replies does not need the ability to send them.

**Insecure output handling**
Treating AI output as trusted input to other systems: running generated code unreviewed, or piping generated text straight into a database or customer channel. Output is a draft, not a command.

**Supply-chain and poisoning risks**
The model, its plugins and its training data all come from somewhere. Vendor due diligence for AI tools is real due diligence now: what is this built on, what does it call out to, and what happened the last time it failed?

THE ONE-LINE SUMMARY FOR LEADERSHIP

"Just connect the model to everything" is the AI equivalent of giving a new contractor keys to every room on day one. Capability is easy. Containment is the work.

**Who owns it:** security owns the threat model and the connector reviews; whoever deploys an agent owns its permissions.

> KEY IDEA
>
> New names, old discipline: least privilege, verified inputs, reviewed outputs, logged actions.

# 7 Governance

**What control actually looks like in companies that have it. Less paperwork than you fear.**

Governance sounds like committees and binders. Strip the jargon and it is three abilities: you **know** what your company uses AI for, you can **see** it happening, and you can **stop** it when something is wrong. Everything in this chapter is one of those three abilities wearing work clothes.

In companies that govern AI well, you can point at four layers, and each one is something real that someone switched on, not a principle on a poster.

## Layer one: policy people actually read

One page, written in plain language: the approved tools, the staff rules, the red lines. Published where everyone can find it, next to a visible list of approved uses. The visible list matters more than the policy. When people can see what is allowed, they stop improvising in the dark, and shadow AI shrinks on its own.

## Layer two: access that follows the person

AI happens through company accounts, not personal ones. That single move puts AI under the controls you already trust everywhere else: people get access when they join, lose it when they leave, and only the right roles can point AI at confidential or personal data. If staff are logging into private accounts to do company work, layer two does not exist yet.

## Layer three: visibility

This is the layer most companies skip, and the one that separates governance from hope. Enterprise AI tools keep logs: who used what, when, connected to which data. Admin dashboards show usage across teams. Audit trails let you answer, months later, "what happened here?" If a tool cannot show you what your company does with it, that tool is fighting your governance, whatever else it does well. Make visibility a buying requirement, not an afterthought.

## Layer four: a human in the loop

The last layer lives inside each workflow: a named person reviews AI output before it takes effect, and the review is real, not a rubber stamp. Drafts get a reader. Decisions get a decider. Agents that can act get supervision and the minimum permissions their task needs. The higher the stakes, the heavier this layer gets, which is exactly the risk-matching idea from page one.

## And when something goes wrong

Treat AI incidents like security incidents: define them (data in the wrong tool, a bad output that reached a customer, an agent doing something nobody asked), give people a no-blame way to report them, and write down what happened. Then put a review date on every AI use you deploy. Things should be retired when they stop earning their risk, not just launched and forgotten.

**LEADERSHIP QUESTIONS**

- Could we list, today, what our company uses AI for? Could we prove it with logs?
- Is every AI tool in use a company account with admin controls switched on?
- For each important use: who is the human in the loop, by name?
- If a tool misbehaved this morning, who would know, and who could switch it off?

> **KEY IDEA**
>
> Governance is control plus visibility: know what AI is used for, see it happening, and be able to stop it. Choose tools that make those three things possible.

# 8 Training your people

**What AI training actually looks like: for the leader buying it and the staff receiving it.**

Every serious study of workplace AI lands on the same finding: the companies getting value are not the ones with the cleverest model. They are the ones whose people were taught to use it. Capability is rented from a vendor. Judgement is built in-house, and the building is cheaper than most leaders expect.

One principle sits above all the formats: **hands-on beats everything**. People do not learn AI from a slide deck about AI. They learn it by using an approved tool on real, low-risk work. That is why training and rollout are the same project: approve the tools, then train people on those tools, on their own tasks, while the stakes are low. Adoption follows use, not memos.

## The training menu

In the real world, workplace AI training comes in levels. Most companies end up mixing all four.

- **Short videos and examples, rolled out to everyone.** Five-minute clips showing how your own teams use the approved tools: how sales drafts a proposal, how support summarises a case. Cheap, fast, and it makes AI feel normal rather than exotic.
- **Quick tutorials on the fundamentals.** How the tools work, what they get wrong, how to brief them properly, when to verify. An hour of honest fundamentals prevents most of the incidents in this book.
- **Structured courses with certificates.** For depth and for tracking. A certificate is not decoration: it lets a leader see, per team, who has actually done the training, and it gives staff a credential worth finishing. Free options exist, including ours.
- **Role deep-dives.** Managers learn to pick use cases, design review into workflows, and measure quality rather than speed. Technical staff take the security chapter as working discipline. Leaders learn the governance and legal landscape well enough to challenge a vendor politely and precisely.

## If you are the CEO

Start earlier than feels necessary. AI is still digestible right now; teams that build habits early compound the advantage, and teams that wait inherit other people's habits from the internet. Budget half a day per person for fundamentals, refreshed yearly. Track completion, celebrate good use publicly, and make sure managers use the tools themselves, because teams copy what managers do, not what policy says.

## If you are staff

Expect to learn what the tools can and cannot do, and to practise on your own work. The mindset that carries you is in Chapter 1: treat AI like a brilliant new intern. Brief it properly, check its work, never let it sign anything, and do not overtrust it just because it sounds

confident. Asking questions and escalating oddities is what good use looks like, not a sign you are behind.

**CULTURE, IN ONE DECISION**

Decide out loud whether using AI well is rewarded or quietly suspicious in your company. If people fear looking lazy for using it, you get hidden use. If they fear nothing at all, you get unreviewed use. The healthy middle is public: use it, verify it, say you used it.

**Who owns it:** the executive sponsor funds it; HR or L&D runs it; managers model it.

> **KEY IDEA**
>
> Train early, train hands-on, and track it. Videos make AI normal, fundamentals make it safe, and certificates make progress visible.

# 9 The law, briefly

**A plain map of the landscape, written from South Africa. Not legal advice, and dated July 2026.**

The most useful legal fact about AI is that **it is not a law-free zone and never was**. Long before any AI-specific act reaches you, your existing obligations already apply: privacy law follows personal data into AI tools, employment law follows AI-assisted decisions about people, consumer protection law follows AI claims in your marketing, and the confidentiality clauses in your contracts follow whatever your staff paste. Map your AI uses to the obligations you already have. That one exercise catches most problems before any new law does.

## At home: South Africa

South Africa has no AI-specific law in force yet, though national AI policy work is under way. What binds you today is the law you already know:

- **POPIA.** Typing personal information into an AI tool is processing personal information, and POPIA applies exactly as it would anywhere else. You stay responsible when a vendor processes data on your behalf, and the cross-border rules matter because most AI tools process data outside South Africa. If personal information flows into a tool, someone needs to have checked where that tool sends it. The Information Regulator is active and expects exactly this kind of homework.
- **Employment law.** Labour law and the Employment Equity Act do not care whether a person or a model shortlisted the candidate. An unfair or discriminatory outcome is unfair and discriminatory either way, which is why Chapter 4 keeps hiring behind elevated review.
- **Governance duties.** King IV already expects boards to govern technology and information responsibly. AI does not need its own chapter in the code to land on the board agenda; it is technology risk with excellent PR.

## Abroad: a taste of what is coming

- **The EU AI Act** is the world's most complete AI law, phasing in through 2025 and 2026, and its core idea matches this handbook: obligations scale with risk. It treats hiring and worker-management systems as high-risk and expects companies to build AI literacy in staff. Even without EU operations, expect its thinking to reach you through client contracts and vendor terms, the same way GDPR arrived years before anyone local asked for it.
- **The United States** has no single federal AI law. Existing regulators apply existing law: exaggerated AI claims and careless data practices are already enforcement territory, AI-assisted hiring already answers to anti-discrimination law, and state-level AI acts are appearing. The direction of travel is the same risk-based instinct.
- **Standards** travel faster than laws. NIST's AI Risk Management Framework and ISO/IEC 42001 give you the vocabulary auditors and enterprise customers increasingly use. You do not need to certify tomorrow; knowing the names and the shape is usually enough to hold your own in a due-diligence questionnaire.

## Intellectual property, in two sentences

Do not assume AI-generated material is automatically yours to protect, or automatically safe to publish; ownership and copyright treatment vary by jurisdiction and are still being settled. Give externally published AI-assisted work an extra review, and read the vendor's terms on output ownership rather than assuming them.

**A DATED PAGE, ON PURPOSE**

This chapter describes the landscape as of July 2026, seen from South Africa. Laws phase in, guidance updates, and case law is forming. Check dates before relying on specifics, and for anything with real exposure, ask a lawyer rather than a chatbot. That last clause is not a joke; it is Chapter 4.

> **KEY IDEA**
>
> POPIA follows the data, employment law follows the decision, and your contracts follow the paste. New AI law arrives on top of those duties, not instead of them.

# 10 The rollout roadmap

**Seven stages from first principles to scaled value, and the two tool decisions that decide how safe the journey is.**

Everything in this handbook assembles into one sequence. It works for a ten-person firm and a thousand-person one. Only the paperwork scales.

1. **Set the principles.** One page: the core idea from the front of this book, the staff rules, the red lines. Signed by leadership. Sent to everyone.
2. **Approve the tools.** Pick a small set, verify their data handling, switch on the enterprise controls, and make access easy. The approved path must be more convenient than the shadow path, or the shadow path wins.
3. **Train the people.** Hands-on, on the approved tools, before the pilots. Not after the incident.
4. **Run low-risk pilots.** Two or three use cases from Chapter 3. Each gets a named owner, a redesigned workflow and a baseline measurement. Six to eight weeks is enough to learn.
5. **Measure and decide.** Quality and outcomes, not just speed and enthusiasm. Kill what did not work, and do it in public. It buys credibility for what did work.
6. **Harden the controls.** Before scaling: the approval path, the incident path, logging on every connected tool, and review dates on everything deployed.
7. **Scale deliberately.** Function by function, workflow by workflow, each new use through the approval path. This is also when agents deserve a look, with Chapter 6 open on the desk.

## The two tool decisions that matter most

Stage two hides the decisions leaders most often get wrong, so they deserve daylight. Before comparing brands, settle these two things.

**First: demand visibility.** Chapter 7 said governance is control plus visibility. That is a buying requirement, not a slogan. Whatever you approve must let you see how your company uses it: admin controls, usage logs, an answer to "what happened here?" months later. A tool that keeps you blind fights your governance every day you run it.

**Second: separate confidential work from public work.** Sort your AI work into two lanes before anyone picks a product.

- The **public lane** is work you could comfortably do in a coffee shop: drafts, brainstorms, summaries of material that is not sensitive. Mainstream assistants, on enterprise accounts, serve this lane well.
- The **confidential lane** is client files, contracts, financials, personal records: the work your business actually runs on. This lane has harder requirements. The data cannot leave an environment you control, which rules out public chatbots however good they are. Answers must be grounded in your own documents and show their sources, so people can verify instead of trust. And every question and answer must be logged, because POPIA and your clients will eventually ask.

Most companies discover the second lane late, usually the day someone pastes a client contract into a free chatbot. The better order is to know the requirements before that day: private deployment, your own documents, visible sources, real logs. Tools built for the confidential lane exist. The point of this chapter is that you should be able to name the requirements before any vendor, including us, shows you a demo.

**THE BLOCKERS TO EXPECT**

They are rarely technical: unclear ownership, no training budget, tools approved but data rules unwritten, leadership excited but unaligned, and workflows never redesigned, so AI stays a novelty layered on old processes. Every one of these is cheaper to fix in stage one than in stage seven.

> **KEY IDEA**
>
> Principles, tools, training, pilots, measurement, controls, scale. In that order. And two tool rules above all: never approve what you cannot see, and never let confidential work into tools you do not control.

KEEP THIS PAGE

# The five decision rules

Almost every "can we use AI for this?" question resolves against these five rules. They are the handbook, compressed.

1. If the task is low-risk and easy to review, AI can usually assist. Start here.
2. If the task involves confidential, regulated or personal data, use only approved enterprise tools through approved pathways.
3. If the task affects a person's employment, legal rights, finances, health or safety, require elevated review and documented accountability.
4. If the tool can take actions rather than just draft, treat it as a higher risk class than a chat assistant, with least privilege and logs.
5. If nobody can say who owns the outcome, the workflow is not ready. Stop and name an owner.

## A compact glossary

**Acceptable use policy**
The written rules for how staff may use AI tools. The staff rules page in this book is a starting template.

**Audit log**
The record of who used a tool for what. The thing you will wish existed after an incident, so switch it on before one.

**Data classification**
Sorting information by sensitivity (public, internal, confidential, personal) so rules can follow the data.

**Impact assessment**
A structured look, before deployment, at who a system could affect and how it could fail. Required in some jurisdictions for high-risk uses; wise everywhere.

**Least privilege**
Granting a system or person only the access their task requires. The first principle of agent safety.

**Red teaming**

Deliberately attacking your own AI setup (prompt injection, data extraction, misuse) to find failures before others do.

**Retention**
How long a tool keeps your conversations and files. An enterprise setting worth verifying, not assuming.

**Vendor due diligence**
Checking what an AI product is built on, what it does with your data, and what its terms actually promise, before it touches real work.

THE SENDABLE ONE-PAGER

# AI at work: the short version

This page stands alone. Send it to every member of staff as-is, or adapt it into your acceptable use policy.

DO

- Use the tools the company has approved.
- Give the AI proper context: goal, background, example, format.
- Verify important output against a source before acting on it.
- Use AI to draft, summarise, organise and explore options.
- Say when work was AI-assisted, where policy requires it.
- Report anything strange: wrong output that shipped, data in the wrong place, a tool acting oddly.

DON'T

- Paste confidential, client or personal data into unapproved tools.
- Let AI make final calls about people, money, legal, medical or safety matters.
- Send AI output you have not read to anyone who matters.
- Trust confident tone, precise-looking numbers or tidy citations without checking.
- Connect AI tools to company systems without IT's sign-off.
- Assume a free tool is private. Assume the opposite.

> AND THE SENTENCE TO REMEMBER
>
> AI drafts. People decide. Owners are named. Everything important gets verified.