A PRACTICAL HANDBOOK FOR UK BUSINESSES THAT HANDLE PERSONAL DATA

Your Data, Your Rules

AI and privacy in plain English for UK small and medium-sized businesses. What happens to what you type, and what to do about it.

AUTHOR
Ted Milton
Founder, Expedient AI

EDITION
July 2026
Version 1.0

FORMAT
A4 print edition
10 pages, worksheets included

Expedient AI Ltd

INDEPENDENT · BRISTOL · UK

1 THE ACTUAL QUESTION

Most guides say "be careful with data" and move on. This one does not.

The question is not "is this safe?" The question is "what happens to what I type?"

If you read most AI guides, you will find a paragraph that says something like "always be careful with sensitive data" and a link to the vendor's privacy page. That is not useful. It does not tell you what happens to the email you pasted in, who can read it, how long it is kept, or whether it will be used to train the next model.

The real question is more specific. It has four parts.

What happens to what I type?

Does it become a prompt that is processed and discarded, or does it enter a training pipeline?

Who can see it?

The model, the vendor's staff, third parties, or a court order?

How long is it kept?

For the duration of the request, thirty days, forever, or until you ask for it back?

What evidence do I have?

A written contract, a checkbox in a console, or a sales call I had last Tuesday?

These are not exotic questions. They are the same questions you would ask any other supplier who handled personal data on your behalf. AI vendors sometimes answer them well, sometimes answer them vaguely, and sometimes cannot answer them at all. Each of those responses tells you something.

> **The rule for this guide.** If a vendor cannot answer these four questions in writing, in plain English, before you sign, you do not yet have a vendor. You have a sales call.

What this guide covers

The four data tiers used by most UK organisations. What actually happens when you press enter on a prompt. The matrix that decides what is safe to send where. Anonymisation, and why removing a name is not enough. What the ICO expects under UK GDPR. The EU AI Act, in the parts that affect an SME using ChatGPT or Copilot rather than a provider building models.

2 CLASSIFY BEFORE YOU SEND

Four tiers. One rule.

Classify the input. Apply the highest tier to the whole input.

Before you decide where data can go, you have to decide what kind of data it is. Most UK organisations, even ones without a formal data protection programme, can sort information into four tiers. The classification is not a legal document. It is a working tool that prevents accidental disclosure.

TIER	EXAMPLES	WHY IT MATTERS
Public	Marketing copy, published prices, press releases.	Safe to use anywhere. No expectation of confidentiality.
Internal	Process notes, training manuals, internal FAQs, anonymised benchmarks.	Embarrassing if leaked, not harmful. Manageable with standard controls.
Confidential	Customer details, contract terms, financial figures, employee records.	Harmful if leaked. Requires an approved environment and a written basis.
Restricted	Special category data under UK GDPR: health, ethnicity, sexual orientation, criminal records, financial vulnerability.	Harmful and regulated. Default is no entry to AI without explicit, documented basis.

The rule

Apply the highest tier to the whole input, not to each piece of data within it. If you paste one customer email into a prompt to rewrite it, the entire prompt is Confidential, even if the prompt itself is a sentence. The model receives the whole input. Its memory does not segment.

> **EASY MISTAKE** Treating a prompt as lower tier than the data inside it. "Summarise this" is a Public instruction. The customer record it summarises is Confidential. The whole conversation is Confidential.

3 FROM KEYSTROKE TO MODEL

What happens between you and the model.

Training, inference, retention, who sees your prompt. And what "we don't train on your data" actually means.

Every prompt you send passes through four stages. Each stage has different rules, different people, and different retention. Most vendor pages blur these stages together. They are not the same.

STAGE	WHAT HAPPENS	WHAT TO CHECK
Transport	Your prompt travels over the internet to the vendor's servers.	TLS in transit, at minimum. Most vendors meet this. None is a differentiator.
Inference	The model produces a response. This is what you see on screen.	Whether the inference runs on shared infrastructure or a private deployment.
Retention	Your prompt and the response are stored for a period. Days, weeks, months or indefinitely.	Retention period, deletion on request, geographic location of storage.
Training	The prompt may be used as training data for the next model version.	Whether you opted in, opted out, or have no choice.

The two tiers of service

Most major vendors offer a free or consumer tier, and an enterprise tier. The differences that matter are not the model. They are the controls around the model.

Consumer / free tier. Your data is typically used to improve the service. Retention is longer. Reviewers may read conversations. You do not get a Data Processing Agreement. Geographic location is wherever the vendor's free-tier infrastructure runs.

Enterprise / business tier. By default, prompts are not used for training in most major vendors. Shorter retention. A signed DPA. Often a regional data residency option. Costs more. Worth it for any business data above the Internal tier.

Reading "we don't train on your data"

This phrase, taken on its own, is incomplete. It does not say who can read the data. It does not say how long it is kept. It does not say whether aggregated patterns are extracted. It does not say what happens on contract termination. Ask the follow-up questions. The honest vendor answers them in writing.

> **The test.** Can the vendor give you a written, signed commitment covering inference, retention, training, geographic location and reviewer access, for the specific product and tier you are buying? If not, the phrase on the website is not the same as the contract.

4 WHERE EACH TIER CAN GO

The safe-use matrix.

Data tier across the top. Environment down the side. What is allowed, and what controls are needed.

Two questions decide whether a particular use is safe. What tier is the data? What environment are you sending it to? The matrix below sets out, for the most common cases, what is allowed with no controls, what is allowed with controls, and what should not be done at all.

DATA TIER	UNAPPROVED PUBLIC AI (FREE CHATGPT, PUBLIC COPILOT, ETC.)	APPROVED BUSINESS ENVIRONMENT (ENTERPRISE TIER, PRIVATE DEPLOYMENT, INTERNAL TOOL)
Public	Allowed. No controls needed.	Allowed. No controls needed.
Internal	Discouraged. Strip names and identifiers before sending.	Allowed. Standard acceptable-use policy.
Confidential	Not allowed. Even redacted data may contain combinations that re-identify.	Allowed under a signed DPA, with logging, review and access controls.
Restricted	Not allowed.	Allowed only with documented lawful basis, DPIA completed, and review by a named owner.

What controls are needed for Confidential data

Approved environment is not enough on its own. The minimum controls for Confidential data in AI are these.

- A signed Data Processing Agreement with the AI vendor.
- Access controls that match who is allowed to see the data normally.
- Prompt and response logging, retained per your retention policy.
- A named human reviewer for any output that leaves the building.
- An exit plan. How you extract your prompts and history if you leave.

How to apply this on Monday

Print the matrix. Pin it next to the team that uses AI most. Spend fifteen minutes with each team agreeing which cells apply to their work. The conversation is the point. People learn the tiers by arguing about which one a particular dataset belongs to.

5 THE RE-IDENTIFICATION PROBLEM

Anonymisation is harder than you think.

Removing the name does not make the data anonymous.

It feels intuitive. Strip the name. Strip the address. Strip the obvious identifiers. Send the rest to the AI. It is also mostly wrong. Anonymisation under UK GDPR requires that re-identification by any reasonably likely means is not possible. Combinations of ordinary detail often re-identify people in surprising ways.

How re-identification works

The classic example: a hospital discharge dataset in which names had been removed, but which contained date of birth, sex and postcode. The researcher was able to re-identify most patients by linking the dataset to local newspaper reports and a voter roll. The data had been pseudonymised, not anonymised.

The same pattern appears in commercial data. Job title, employer, age band, region and a project description will often identify a specific individual to a competitor. The model does not need a name to produce a useful profile, but the input still contains personal data until it is genuinely anonymised.

> LEGAL TEST Under UK GDPR, data is anonymised only when re-identification is not possible by any means "reasonably likely to be used" by the controller or another party. The test is not "we removed the obvious fields". It is "would a competent person with access to other data be able to identify the individual".

Synthetic data as an alternative

Where you need the shape of real data without the individuals, synthetic data is often a better answer than attempted anonymisation. Synthetic data is generated to mimic the patterns of the real data, with no one-to-one mapping back to a real person. It works well for testing, training and demonstrations. It does not work for any analysis where the specific individual matters.

> **Practical rule.** For Internal-tier use, careful redaction may be enough. For Confidential, treat redaction as a harm-reduction step, not a basis for sending the data to AI. The basis must be the environment, not the redaction.

6 UK GDPR, THE DPA 2018 AND YOU

What the ICO actually says.

UK GDPR and the Data Protection Act 2018 as the floor. The eight questions to answer before putting personal data into AI.

The Information Commissioner's Office regulates data protection in the UK. For an SME using AI, the relevant framework is UK GDPR plus the Data Protection Act 2018. The ICO has published guidance on AI and data protection, and the guidance is updated as practice evolves. Treat it as the floor, not the ceiling.

The eight questions

Before you put any personal data into an AI system, answer these questions in writing. If you cannot, the data does not go in.

1. What is the lawful basis for processing this data under UK GDPR Article 6?
2. If the data is special category, what is the Article 9 condition?
3. Has a Data Protection Impact Assessment been completed where required?
4. Is the AI vendor a controller or a processor, and do you have a signed contract reflecting that?
5. Where is the data stored, and does that location meet your transfer rules?
6. How long is the data retained by the vendor, and how do you request deletion?
7. Who at the vendor can see the data, and under what circumstances?
8. What is your process if the vendor suffers an incident involving your data?

> **WHEN A DPIA IS REQUIRED** The ICO treats AI processing that is "likely to result in a high risk to the rights and freedoms of individuals" as requiring a DPIA. Systematic and extensive evaluation of personal aspects, including profiling, is named as a likely trigger. When in doubt, do the DPIA. It is a working document, not a paperwork exercise.

What the ICO does not yet require

The ICO has been clear that it does not intend to ban generative AI. It expects organisations to apply existing data protection law sensibly and to be able to evidence their decisions. Good faith effort, written down, has so far been treated more favourably than rushed adoption without consideration.

This guide is not legal advice. For questions involving special category data, large-scale monitoring or significant decisions about individuals, speak to a data protection specialist.

7 WHAT MATTERS FOR AN SME, NOT A PROVIDER

The EU AI Act in 90 seconds.

For an SME that uses ChatGPT or Copilot, not for a company that builds models.

The EU AI Act is in force. The obligations arrive in waves between 2025 and 2027. Most of the headline provisions concern providers of high-risk AI systems, not the small business that uses ChatGPT to draft a customer reply. This page is about what matters for the user, not the provider.

The dates that matter

FROM	OBLIGATION	WHY IT MATTERS TO YOU
August 2025	Obligations on providers of general-purpose AI models (GPAI).	The provider of the model you use has new documentation duties. You may benefit from better model cards.
August 2026	Transparency obligations for AI systems that interact with people or generate content.	Outputs of AI used to inform the public may need to be labelled as such.
August 2027	Obligations for high-risk AI systems in Annex III, including those used for employment, credit and biometric identification.	If you deploy AI in HR, credit or similar decisions, the obligations land here.

What matters for an SME using ChatGPT or Copilot

Three things. The first is transparency to the people affected. If a customer or employee is reading text an AI drafted for them, in some contexts they need to be told. The second is keeping records of high-risk uses, even where you are a user rather than a provider. The third is choosing providers that can give you the documentation the Act requires them to produce.

What matters for an AI provider

Almost everything else. Conformity assessment, risk management, post-market monitoring, registration of the system in an EU database. If you are building AI for sale in the EU, this guide is not for you. Speak to a specialist.

Northern Ireland and the UK

The UK is not applying the EU AI Act directly. Northern Ireland has a particular position under the Windsor Framework, which has practical consequences for businesses operating on both sides of the border. Check current guidance if this affects you.

> **Bottom line.** Most SMEs using consumer or enterprise AI tools do not need to do anything dramatic under the EU AI Act in 2026. Apply the seven-tier system, get the contracts in writing, and revisit this page annually.

8 THE MONDAY MORNING VERSION

A 30-minute data audit.

Find out, in writing, where your data is going.

The single most useful exercise is a 30-minute audit of where your data actually goes today. Not where the policy says it goes. Where it actually goes. Most organisations find at least one surprising use, often an external AI tool used by an individual team member with no formal approval.

How to run the audit

List the AI tools in use. Free, paid, browser extensions, automations. Ask each team in turn.

For each tool, list the data tier sent. Public, Internal, Confidential or Restricted.

For each row, ask the four questions. Training, retention, reviewer access, geographic location.

Mark the rows that fail. Confidential data sent to an unapproved environment is the most common failure.

Decide what to fix this month. Not everything at once. One row, one fix.

What to do about the surprising rows

Be calm. People rarely set out to leak data. They use a tool because it is faster. The fix is usually one of three.

Move it to an approved environment. If the use is reasonable, give the team the enterprise tier and the documented basis for processing.

Reduce the data tier. Strip identifiers, use synthetic data, change the workflow so less personal data is needed.

THE AUDIT SUMMARY TO WRITE DOWN

Total tools in use. Tools sending Confidential or Restricted data. Tools without a signed DPA. Tools where the training, retention, reviewer access and location questions cannot be answered. The single fix chosen for this month.

This guide is general information, not legal advice. For questions involving special category data, large-scale monitoring, or significant decisions about individuals, speak to a data protection specialist. Vendor privacy pages and ICO guidance change. Check the current position before relying on any figure in this PDF.

09 COLOPHON

The series and how to reach us.

Six field guides for UK SMEs. One Bristol-based company behind them.

The series

Six field guides for UK small and medium-sized businesses. Independent, plain-English, written by a practitioner.

01 · Where AI Actually Helps
Choose one worthwhile use case, scope a 30-day pilot and estimate the return before buying software.

02 · The 30-Day AI Pilot
A day-by-day guide to running the pilot you scoped in edition 01.

03 · Your Data, Your Rules
AI and privacy in plain English for UK businesses handling personal data.

04 · Writing AI Instructions That Actually Work
The six-part structure that turns vague prompts into reliable outputs.

05 · The AI Vendor Interrogation Kit
Eight questions to ask before signing a contract with any AI vendor.

06 · Your First AI Employee
A non-technical guide to autonomous AI agents for small businesses.

Colophon

PUBLICATION

Your Data, Your Rules

July 2026 · Version 1.0

Author: Ted Milton, Expedient AI Ltd

CONTACT

Expedient AI Ltd

expedientai.co.uk

> You can use this guide without contacting Expedient AI. If you would value an independent review of your data handling or AI controls, Expedient AI Ltd offers practical AI implementation support for UK SMEs.