Founders AI
founders-ai.dev

PRACTICAL AI FIELD GUIDE

AI Privacy System

What Not to Put Into AI

A practical privacy guide for founders, business owners, creators, consultants, and professionals who want to use AI without leaking sensitive information.

- Classify data
- Redact context
- Check tools
- Limit access
- Avoid secrets
- Log mistakes

AI gets more useful when you give it context. That is also where people get careless.

They paste customer emails. They upload contracts. They drop private screenshots. They copy employee notes, API keys, financial data, medical details, invoices, call transcripts, code, and internal documents into whatever AI tool is closest.

Sometimes that is fine. Sometimes it is not. The difference is not whether AI is “safe” in general. The difference is whether the data, the tool, the settings, and the use case match.

> **The privacy rule is simple: do not give AI data it does not need, and do not give sensitive data to a tool you have not checked.**

This guide gives you a practical system for deciding what can go into AI, what should be redacted first, and what should stay out.

Why this matters

AI tools can feel like a private conversation, but they are still software systems with data policies, retention rules, admin settings, third-party integrations, logs, model features, and human or automated review processes. OWASP warns that LLM applications can expose personal information, financial details, health records, confidential business data, security credentials, legal documents, proprietary algorithms, source code, and another user’s personal data (OWASP LLM02:2025 Sensitive Information Disclosure).

NIST also warns that generative AI can create privacy risks through leakage, unauthorized disclosure, de-anonymization, memorization, and inference of sensitive information (NIST AI 600-1). In plain English, AI privacy risk is not only “will this company train on my prompt?” It is also “did I paste information I never needed to paste?”

The goal is not paranoia. The goal is a habit:

> **Classify first. Redact second. Prompt third.**

The four data classes

Before you use AI, classify the data. This takes less than a minute and prevents most mistakes.

Class	What it means	AI use rule
Public	Already public, approved, or safe to share	Usually safe
Internal	Not secret, but meant for your team	Use with care
Confidential	Customer data, strategy, finances, private docs, contracts	Redact or use approved business tools
Restricted	Secrets, passwords, keys, medical/legal/financial identifiers, regulated data	Do not paste unless you have the right tool, agreement, and approval

If you are unsure, move the data up one class. Treat “maybe confidential” as confidential.

The never-paste list

Do not paste these into a general AI chat unless you have verified the tool, settings, retention policy, permissions, and legal/business approval:

- API keys, passwords, tokens, private keys, recovery codes, or connection strings
- Full customer records, especially names plus emails, addresses, payment details, health data, or account details
- Employee records, HR notes, compensation data, performance notes, or private messages
- Bank statements, tax records, credit card details, invoices with full identifiers, or financial account data
- Legal documents that include private client facts, case strategy, signatures, or protected information
- Medical records, diagnoses, prescriptions, lab results, insurance data, or patient identifiers
- Proprietary code, private repositories, unreleased product plans, or trade secrets
- Private screenshots that reveal tabs, names, emails, account IDs, tools, browser sessions, or internal URLs
- Anything you would be embarrassed, harmed, or legally exposed by if it appeared in the wrong place

OWASP specifically calls out credentials, legal documents, confidential business information, source code, and personal data as sensitive information risks for LLMs (OWASP LLM02:2025 Sensitive Information Disclosure).

The privacy decision tree

Use this before uploading, pasting, or connecting data:

1. **Is the data public?** If yes, it is usually safe.
2. **Does the AI need the exact data?** If not, summarize or anonymize it.
3. **Does the data identify a person, customer, employee, patient, account, or private business matter?** If yes, treat it as confidential.
4. **Does the data include secrets, credentials, regulated information, or legal/medical/financial identifiers?** If yes, do not paste it into a normal chat.
5. **Is the tool a personal consumer tool or a business/enterprise tool?** Use the safer tool for sensitive data.
6. **Does the tool use your data for training by default?** Check the settings or policy.
7. **Does the tool retain files, chats, logs, or connected-tool data?** Check before uploading.
8. **Are third-party connectors involved?** Check the connector's policy too.
9. **Would you be comfortable with an admin, vendor, auditor, or future teammate seeing this prompt?** If not, redact first.

If the answer is still unclear, do not paste the raw data. Create a sanitized version.

The safer context method

AI often does not need raw private data. It needs the structure of the problem.

Instead of pasting this:

```
John Smith at ACME Roofing owes $14,822.40, has been late 47 days, lives at [address],
and said in an email that his wife lost her job. Write a collection email.
```

Paste this:

```
A customer is 47 days late on a $14.8k invoice. They shared a personal hardship.
Write a firm but respectful follow-up that offers a payment-plan option without
mentioning private personal details.
```

The second prompt gives AI enough context to help without exposing unnecessary personal information.

The redaction workflow

Use this workflow before sharing confidential context with AI:

1. **Remove identifiers**: Names, emails, phone numbers, addresses, account numbers, IDs.
2. **Remove secrets**: Keys, tokens, passwords, private links, internal URLs.
3. **Generalize numbers**: Use ranges or rounded values unless exact numbers matter.
4. **Replace people with roles**: “Customer A,” “employee,” “vendor,” “founder,” “manager.”
5. **Replace companies if needed**: “A B2B SaaS company,” “a roofing contractor,” “a local clinic.”
6. **Keep only the useful facts**: Delete anything the AI does not need.
7. **Check screenshots**: Crop browser tabs, sidebars, account menus, email addresses, and filenames.

The May 2025 joint AI data security guidance recommends data minimization, anonymization, secure transfer, data classification, and access controls across the AI lifecycle (Joint Cybersecurity Information Sheet). The everyday version is: only share the minimum context required.

The redaction prompt

Use AI to help prepare a safer version, but do not paste the sensitive version into the wrong tool. If the data is sensitive, redact manually first or use an approved secure tool.

Use this prompt on already-safe text:

```
Review the text below and create a privacy-safe version.

Remove or replace:
- Names
- Emails
- Phone numbers
- Addresses
- Account numbers
- API keys, tokens, passwords, private URLs, and credentials
- Company secrets
- Private customer, employee, medical, legal, or financial details

Keep only the minimum context needed for the AI task.
Return:
1. Redacted version
2. What was removed
3. Any remaining privacy risks

Text:
[paste already-safe or partially redacted text]
```

Important: do not use the redaction prompt as an excuse to paste highly sensitive raw data into a random AI tool. Redaction is a process. The tool choice still matters.

Check the tool, not just the model

Privacy depends on the product, plan, settings, and feature, not just the model name. A consumer chat app, business workspace, API, browser extension, file-upload feature, code tool, and third-party connector can all handle data differently.

Microsoft says Microsoft 365 Copilot respects existing identity models and permissions, inherits sensitivity labels, applies retention policies, supports audit of interactions, and does not use prompts, responses, or Microsoft Graph data to train foundation models (Microsoft Learn). Google says Gemini in Workspace can only retrieve data the user has permission to access and does not use Workspace customer data to train or improve underlying Gemini/Search models outside Workspace without permission (Google Workspace).

That does not mean every AI tool has the same defaults. Anthropic's API data-retention page distinguishes zero data retention eligible API features from features that retain data, and it warns that third-party integrations are not covered by Anthropic's ZDR arrangement (Anthropic Claude docs).

Before using sensitive data, check:

- Is this a consumer account, business account, enterprise account, or API?
- Is my data used for model training by default?
- Can training be disabled?
- How long are chats, files, logs, and outputs retained?
- Are uploads stored until deleted?
- Do admins have access to prompts or logs?
- Are third-party tools or connectors involved?
- Does the tool inherit my existing file permissions?
- Does it have audit logs, DLP, retention controls, or zero data retention options?

The permission trap

AI connected to your workspace can only be as safe as your permissions. If a file is shared too broadly, AI may be able to retrieve it for anyone who already has access.

This is why permissions matter before AI rollout. Microsoft says Copilot respects identity models and permissions, while Google says Gemini can only retrieve data the user has permission to access (Microsoft Learn, Google Workspace). That is good, but it also means old folder mistakes can become AI-searchable mistakes.

Before connecting AI to email, docs, drives, CRMs, or project tools:

- Review shared folders.
- Remove old contractors and former employees.
- Check “anyone with the link” files.
- Separate HR, finance, legal, medical, and customer data.
- Use least privilege: people should only access what they need.
- Apply labels where your tool supports them.
- Turn on audit logs where available.

The privacy system is not only about prompts. It is also about access.

Do not hide secrets in prompts

A prompt is not a vault. A system prompt is not a security boundary.

OWASP’s system prompt leakage guidance warns that system prompts can inadvertently contain secrets and says system prompts should not be considered secrets or used as security controls (OWASP LLM07:2025 System Prompt Leakage).

Never put these in system prompts, custom instructions, bot descriptions, workflow descriptions, or agent memory:

- API keys
- Auth tokens
- Database credentials
- Passwords
- Private URLs
- Internal system architecture
- User roles and permission structures
- Customer secrets
- “Hidden” business rules that would create risk if discovered

If a tool needs a secret, store the secret in a secure credential manager or backend system. Let the AI request an action. Let deterministic software decide whether the action is allowed.

The everyday AI privacy policy

If you are starting from zero, use this policy:

```
AI Privacy Policy

1. Public information can be used freely.
2. Internal information can be used if it does not expose customers, employees, credentials, or private strategy.
3. Confidential information must be redacted before use unless an approved business tool is being used.
4. Restricted information cannot be pasted into AI without explicit approval.
5. Secrets, keys, passwords, tokens, and private credentials are never pasted into AI.
6. Customer, employee, medical, legal, and financial data must be minimized or anonymized.
7. Files and screenshots must be reviewed before upload.
8. Third-party connectors must be checked before use.
9. AI outputs that include private data must not be copied into public places.
10. Mistakes must be logged so the policy improves.
```

Keep this policy simple enough that you will actually follow it.

The 10-minute privacy check

Before using AI for real work, answer these:

- What data am I about to share?
- What class is it: public, internal, confidential, or restricted?
- Does AI need the raw data or just the situation?
- Can I remove names, emails, IDs, account numbers, and secrets?
- Does this tool use my data for training?
- How long does this tool retain chats or files?
- Are third-party connectors involved?
- Who else can access this file, prompt, output, or workspace?
- What is the worst reasonable outcome if this leaks?
- Should I use a safer tool, redact more, or not use AI here?

If you cannot answer those questions, pause.

The safe-use examples

Use case	Bad prompt	Better prompt
Customer email	Paste full private customer thread	Summarize the issue without names or identifiers
Legal review	Paste a signed private contract into a consumer tool	Use an approved secure tool or paste only redacted clauses
Code help	Paste repo with tokens and private URLs	Remove secrets and share the smallest failing snippet
Hiring	Paste candidate names and private notes	Use anonymized candidate profiles and objective criteria
Finance	Paste full bank statement	Share categorized totals or fake sample rows
Medical	Paste patient details	Do not use unless the tool and workflow are approved for PHI

The privacy habit is not “never use AI.” The habit is “share the minimum useful version.”

What to do if you already pasted something sensitive

Do not panic. Do this:

1. Delete the chat or file if the tool allows it.
2. Check whether the tool retains deleted content.
3. Rotate any exposed passwords, API keys, tokens, or private links.
4. Tell the relevant owner if customer, employee, legal, medical, or financial data was exposed.
5. Review the tool’s data policy.
6. Update your AI privacy policy so it does not happen again.
7. If this happened in a business, log it as an incident and ask for professional guidance if needed.

The FTC has warned that AI companies must honor privacy and confidentiality commitments, and that there is no AI exemption from existing laws (Federal Trade Commission). For a user or business, the matching lesson is: privacy promises, tool settings, and data handling matter.

The Founders AI takeaway

AI is not dangerous because it reads context. AI becomes risky when people provide context without thinking.

You do not need to be a privacy lawyer to use AI responsibly. You need a simple system:

- Classify the data.
- Remove what the AI does not need.
- Use the right tool for the sensitivity level.
- Check permissions before connecting AI to workspaces.
- Never put secrets in prompts.
- Log mistakes and improve the policy.

That is enough to use AI more confidently without treating every prompt like a private vault.