

PROMPT HYGIENE GUIDE

Practical Workplace Rules for Safe AI Use



Version 1.0 – August 2025



PROMPT HYGIENE: KEEPING SENSITIVE DATA OUT OF AI

ONE BAD PROMPT CAN LEAK MORE THAN YOU THINK.

Even the most secure AI deployment can't protect you if you feed it sensitive, confidential, or regulated information it was never meant to process. This quick-reference guide helps you decide what's safe to share, what needs checking first, and what should never go into an AI system, whether public, enterprise, or fully private.



PROMPT SAFETY ZONES

Zone	What It Means	Examples	Action
Green Zone	Information that is safe to use because it's public, non-sensitive, and non-identifiable.	Published marketing content, FAQs, product specs already on your website.	Proceed : still check for accuracy before sharing outputs.
Yellow Zone	Information that may be safe but depends on context, contractual obligations, or consent.	Draft policy documents, internal process descriptions, generic training datasets.	Pause: confirm with Legal/DPO/ Manager before use.
Red Zone	Information that is regulated, confidential, or would harm the organisation if disclosed.	Customer PII, health records, financial statements, trade secrets, unreleased product details.	Never prompt: find or create anonymised/ synthetic alternatives.

WHY THIS MATTERS FOR ARTICLE 4 AI LITERACY

Under Article 4 of the EU AI Act, organisations must ensure staff have the knowledge and skills to use AI responsibly, including awareness of information security risks. Recognising and respecting these “safety zones” is a core part of that literacy, protecting both the organisation and individuals from breaches, fines, and reputational harm.





Prompt Hygiene: Keeping Sensitive Data Out of AI

THE RED ZONE: WHAT NEVER GOES INTO AI

The Red Zone is information that, if entered into any AI system: public, enterprise, or private, could expose your organisation to regulatory penalties, legal disputes, competitive losses, or reputational harm.

If in doubt, treat it as Red Zone and do not prompt.



REGULATED PERSONAL DATA


Type	Examples
Obvious	Customer names, addresses, phone numbers; Passport or national ID numbers; Bank account or credit card numbers
Less Obvious	Email addresses tied to a customer account; Indirect identifiers (e.g., job title + small town); Metadata in documents or images revealing identity

BUSINESS CONFIDENTIALS

Type	Examples
Obvious	Trade secrets, formulas, proprietary algorithms; Unreleased product designs/specifications; Strategic business plans
Less Obvious	Supplier pricing lists or contract terms; Internal financial forecasts before public release; Unannounced partnership discussions

OPERATIONAL DETAILS

Type	Examples
Obvious	Security system configurations; Access credentials or network diagrams; Incident response playbooks
Less Obvious	Process weaknesses being fixed; Internal audit findings; Staffing rosters linked to specific roles or sites

 **Pro Tip:** Combine Red Zone awareness with (Prompt Hygiene techniques from [AI Literacy – Ethical Prompt Fluency™](#)) to create safe, anonymised, or synthetic versions for training AI without exposing the originals.





Prompt Hygiene: Keeping Sensitive Data Out of AI

THE YELLOW ZONE: CHECK BEFORE YOU PROMPT

The Yellow Zone is information that might be safe to use in AI - but only after confirming context, permissions, and any contractual or compliance obligations.

If you're unsure, pause and check with the right person before using it.



CONTEXT-SENSITIVE INTERNAL DOCUMENTS


Type	Examples
Obvious	Draft policy documents; Internal process descriptions; Non-public training manuals
Less Obvious	Early-stage project notes; Internal meeting minutes; In-progress reports containing speculative figures

CONTRACTUALLY CONTROLLED INFORMATION

Type	Examples
Obvious	Third-party data covered by NDAs; Vendor-supplied materials marked "Confidential"
Less Obvious	Screenshots from supplier portals; Excerpts from draft agreements; Internal performance dashboards sourced from contracted partners

SENSITIVE OPERATIONAL INSIGHTS

Type	Examples
Obvious	Workflow diagrams showing internal systems; Resource allocation charts; Shift rosters without direct identifiers
Less Obvious	Informal process workarounds; Bottlenecks under investigation; Draft security process changes not yet approved

 **Pro Tip:** When in the Yellow Zone, use a quick "Permission + Purpose" check ask:

1. Do I have explicit approval to share this information with an AI system?
2. Is there a business-critical reason to do so, or can I reframe/simplify the prompt to remove sensitive details?





Prompt Hygiene: Keeping Sensitive Data Out of AI

THE GREEN ZONE: SAFE TO USE - WITH BASIC CARE

The Green Zone is information that is already public, non-sensitive, and non-identifiable.

It's generally safe to use in AI prompts, but you should still check for accuracy before acting on the AI's output.



PUBLIC-FACING ORGANISATIONAL CONTENT


Type	Examples
Obvious	Published marketing materials; Public product specifications; Website FAQs
Less Obvious	Public press releases; Company social media posts; Job ads already live on external platforms

INDUSTRY-COMMON KNOWLEDGE

Type	Examples
Obvious	Widely known industry standards; Publicly available regulations; Generic "how-to" processes
Less Obvious	Competitor product features already listed on their websites; Industry benchmark data from public reports

NON-IDENTIFIABLE TRAINING OR PRACTICE DATA

Type	Examples
Obvious	Made-up scenarios for practice; Anonymised case studies; Synthetic datasets created for demos
Less Obvious	Obfuscated internal examples (with names and sensitive figures changed); Mock-ups created for training purposes

 **Pro Tip:** Even in the Green Zone, never skip verification- AI can misinterpret even public data. Treat all outputs as drafts until reviewed for accuracy and tone.





Prompt Hygiene: Keeping Sensitive Data Out of AI

SAFE ALTERNATIVES: TURNING RISKY DATA INTO PROMPT-READY CONTENT

You don't have to abandon valuable AI tasks just because the original information is in the Red or Yellow Zone.

Use the following techniques to transform sensitive inputs into safe, compliant versions that you can work with confidently.



ANONYMIZATION

Definition: Permanently removing or replacing personal identifiers so individuals cannot be re-identified- even by combining with other data.

Before	After
"Prepare a summary of the complaint submitted by John Smith, Account #4728..."	"Prepare a summary of the complaint submitted by a customer regarding late delivery..."

PSEUDONYMISATION

Definition: Replacing personal identifiers with fictitious labels or codes, while keeping a secure mapping key stored separately for legitimate use only.

Before	After
"List the steps taken by Dr. Susan Lee in the patient's treatment plan..."	"List the steps taken by Doctor A in the treatment plan..."

DATA GENERALIZATION

Definition: Reducing the precision of data so it describes groups or ranges, not specific details.

Before	After
"Identify sales patterns for customers in Colombo who spent over Rs. 100,000 last month..."	"Identify sales patterns for customers in major cities who spent over Rs. 100,000 last quarter..."

SYNTHETIC DATA CREATION

Definition: Generating entirely artificial data that mimics the statistical patterns of the original without revealing real information.

Note: Poorly generated synthetic data can still leak patterns - always verify quality before use.

Before	After
"Run the analysis on our actual 2024 client revenue dataset..."	"Run the analysis on this synthetic dataset modelled on 2024 client revenue"





COMMON MISTAKES THAT PUT SENSITIVE DATA AT RISK

Even experienced professionals can slip up.

Avoid these common errors to keep your AI use safe and compliant:



1. Assuming Enterprise AI = No Risk

Mistake: Believing that because you're using a company-hosted or enterprise AI system, any data is safe to input.

Reality: Enterprise systems still require strict data handling — vendor configurations, default logging, and retention policies vary widely and can expose sensitive content later.

2. Copy-Pasting Without Screening

Mistake: Dropping whole sections of documents into AI without checking for Red or Yellow Zone material.

Reality: Even a single paragraph can contain identifiers or confidential details hidden in text, tables, or metadata.

3. Forgetting About Indirect Identifiers

Mistake: Removing names but leaving job titles, small location data, or project details that can still identify individuals or clients.

Reality: Regulators treat indirect identifiers as personal data if they can be linked back to someone.

4. Mixing Data Zones in a Single Prompt

Mistake: Combining safe (Green Zone) and risky (Red/Yellow Zone) content in one AI request.

Reality: Even one Red Zone element makes the entire prompt risky — and it can't be "unshared" once submitted.

5. Not Documenting Safe-Use Practices

Mistake: Using safe alternatives but failing to record how the data was transformed.

Reality: Without documentation, you can't prove compliance under frameworks like the EU AI Act's Article 4.





Prompt Hygiene: Keeping Sensitive Data Out of AI

KEY TAKEAWAYS

Recognise your data zone before prompting

Green = generally safe,

Yellow = check first,

Red = never prompt.

Apply safe alternatives to transform risky inputs into AI-ready content.

Document your process so you can demonstrate compliance at any time.



NEXT STEPS: BUILD YOUR AI LITERACY FURTHER

This guide is part of the [AI Literacy Series](#) from Libra Sentinel.

To go deeper into practical, compliance-ready AI skills:

- [AI Literacy – Technical Foundations for Non-Tech Roles](#) → Understand how AI systems handle your data and the limits of different deployments.
- [AI Literacy – Ethical Prompt Fluency™](#) → Learn safe, auditable prompting methods for regulated environments.
- [Organizational AI Governance](#) → Build end-to-end governance policies that protect your organisation while enabling responsible AI use.

