EMPLOYEE HANDBOOK · SECURE & ETHICAL AI

Responsible AI Use Guidelines

Practical guidance for the secure, transparent and responsible use of AI systems by employees and organisations.

AUDIENCE
All employees, contractors & teams

OWNER
Chief Information Security Officer

COMPANION TO
AI Security Policy

VERSION
1.0 · 2026 (EN)

CLASSIFICATION
Internal

REVIEW
Annual

[ORGANISATION NAME]

SwissCyberAI · Baar / Zug — Switzerland

2026
EN

How to use these guidelines

These guidelines translate our AI Security Policy into everyday, practical advice. They are written for everyone who uses AI at work — not just technical staff.

Artificial intelligence can help you write, research, analyse, code and create faster. Used carelessly, the same tools can leak confidential data, spread errors, or expose the organisation to legal and reputational harm. These guidelines show you how to get the benefits while staying safe, honest and compliant. They apply whenever you use any AI tool for work — whether it is a public chatbot, an AI feature inside an application you already use, or a system we build ourselves.

If you only read one page, read this one.

The ten golden rules of responsible AI use

1. Use only **approved** AI tools for work.
2. Never paste **confidential or personal** data into public AI.
3. Treat every AI output as a **draft**, not a fact.
4. **Verify** facts, figures, quotes and code before you use them.
5. Stay **accountable** — you own what you send, publish or decide.
6. **Disclose** AI use when honesty or the rules require it.
7. Watch for **bias** and unfair or harmful output.
8. Respect **copyright, privacy and confidentiality**.
9. Keep a **human in the loop** for decisions that affect people.
10. **Report** anything that looks wrong, leaked or misused.

These rules are expanded in the sections that follow. When in doubt, pause and ask your manager or the security team before continuing.

1 Why responsible AI matters

AI is now part of daily work, and that creates both opportunity and responsibility. Using AI well is not only about avoiding mistakes — it protects the people we serve, the trust others place in us, and the value of our own work.

- **Data can leak.** Anything typed into an external AI service may be stored or processed by the provider, and in rare cases surfaced to others.
- **AI can be confidently wrong.** Models invent facts, citations and code that look convincing but are false — this is called a "hallucination".
- **AI can be unfair.** Models can reflect bias in their training data and produce discriminatory or skewed output.
- **Accountability stays with you.** AI is a tool. Responsibility for the outcome always remains with the person who used it.

> **THE CORE IDEA**
>
> AI should **augment your judgement, not replace it**. You remain the professional in the loop — the AI is an assistant whose work you check, own and stand behind.

2 The principles behind these guidelines

Six simple principles sit behind everything here. Keep them in mind and most decisions become straightforward.

Principle	In everyday terms
Accountable	You are responsible for what you produce with AI, just as if you had written it yourself.
Transparent	Be honest about when and how AI was used, where honesty or policy requires it.
Secure	Protect data and access. Use approved tools and never leak sensitive information.
Fair	Check for bias; do not use AI to demean, discriminate or mislead.
Human-centred	Keep meaningful human review over decisions that affect people.
Private	Respect the privacy of colleagues, customers and third parties in every prompt.

3 Before you use AI: choose an approved tool

Not all AI tools are equal. Approved tools have been checked for security and data handling; unapproved ("shadow AI") tools have not, and may send your data anywhere.

- Use only tools on the organisation's **[approved AI tools list]**. If a tool is not listed, request approval before using it for work.
- Prefer enterprise or paid business versions, which usually offer stronger privacy terms than free consumer versions.
- Sign in with your organisational account where required, and never share your AI accounts or API keys.
- If you are unsure whether a feature uses AI or where its data goes, ask before entering anything sensitive.

QUICK DECISION — BEFORE YOU PRESS ENTER

Ask yourself: **(1)** Is this an approved tool? **(2)** Would I be comfortable if this input became public? **(3)** Am I about to include personal or confidential data? If the answer to (2) is "no" or (3) is "yes", stop and use an approved, protected tool — or don't include that data at all.

4 Protecting data: the traffic-light rule

The single most common AI mistake is putting the wrong data into a tool. Use this simple traffic-light guide for what you may share. When unsure, treat data as more sensitive, not less.

Signal	Type of information	What to do
GREEN	Public or non-sensitive: published material, general questions, public facts, non-confidential drafts.	Fine to use with approved AI tools.

Signal	Type of information	What to do
AMBER	Internal but not highly sensitive: internal documents, non-personal project details, routine business content.	Use only with approved tools that keep data internal. Do not use public/free services.
RED	Confidential, personal or regulated: customer/ employee personal data, credentials, financials, trade secrets, legal, health, security details.	Do not enter into general AI tools. Use only where explicitly approved for that data class, with safeguards in place.

GOLDEN RULE FOR DATA

If you would not post it publicly or email it to an outside party, do not paste it into a public AI tool. Personal data (names, contacts, IDs, anything identifying a real person) is **RED** by default.

5 Check the output: never trust blindly

AI produces fluent, confident text — including when it is wrong. Treat every output as a first draft to be verified, especially anything you will share, publish, or act on.

5.1 Verify before you rely

- ☐ Facts, dates, names and statistics — confirm against a trusted source.
- ☐ Quotes, citations and references — AI often invents these; check they exist.
- ☐ Code — review, test and scan it before running; never paste secrets into it.
- ☐ Numbers and calculations — re-check the maths yourself.
- ☐ Legal, financial, medical or safety content — have a qualified person review it.

Do

- Use AI to draft, summarise, brainstorm and explain.
- Cross-check important output against reliable sources.
- Ask the AI to show its reasoning or sources, then verify them.
- Keep the final judgement — and edit — yours.

Don't

- Copy-paste AI output into important work unchecked.
- Assume citations or quotes are real.
- Let AI make a consequential decision on its own.
- Publish AI output as verified fact without checking.

6 Be transparent about AI use

Honesty builds trust. Being open about AI use protects you and the organisation, and is sometimes required by policy, contract, law or professional standards.

- **Disclose when it matters.** Tell people when they are interacting with an AI system (e.g. a chatbot) rather than a human, and when AI materially shaped a decision that affects them.
- **Don't misrepresent.** Do not present AI-generated work as if it were independently authored or verified when that would mislead.
- **Follow context-specific rules.** Some clients, regulators or professional bodies require explicit disclosure — check before you assume.
- **Keep a record.** For significant AI-assisted work, note that AI was used and how, so it can be reviewed later.

> **RULE OF THUMB**
>
> If disclosing that you used AI would change how someone reacts to the work, that is usually a sign you **should** disclose it.

7 Everyday security habits

A few simple habits keep AI use secure and prevent it from becoming an attack path.

7.1 Protect access and accounts

- Use strong, unique passwords and multi-factor authentication on AI accounts.
- Never share accounts, API keys or credentials; store keys securely, never in prompts.
- Sign out of shared devices and review connected apps periodically.

7.2 Beware manipulation and fakes

- **Prompt injection.** Content from websites, documents or emails can hide instructions that try to hijack an AI assistant. Be cautious when an AI acts on untrusted content, and review what it does before allowing actions.
- **Deepfakes & AI phishing.** AI makes convincing fake voices, videos and messages. Verify unusual requests — especially for money, credentials or data — through a second, trusted channel.
- **Over-trust.** A confident tone is not proof. Stay sceptical of AI that asks you to bypass a control or reveal something sensitive.

7.3 No shadow AI

Do not use unapproved AI tools, browser extensions or plug-ins for work. If you think a tool would help, ask to have it assessed and approved — that is how it becomes safe to use.

8 Fair, ethical and respectful use

- **Guard against bias.** AI can reproduce stereotypes and unfair patterns. Review output that affects people (hiring, evaluation, customers) for fairness, and don't rely on it uncritically.

- **No harmful content.** Do not use AI to create harassing, deceptive, discriminatory, or otherwise harmful or unlawful material.
- **Respect people.** Do not generate content about identifiable individuals that you would not be comfortable defending, and never to impersonate or demean.
- **Use good judgement.** "The AI suggested it" is never a justification for something you know to be wrong.

9 Intellectual property and confidentiality

- Do not enter third parties' confidential information or others' copyrighted material into AI tools without the right to do so.
- Be aware that AI output may resemble existing works; check before using generated text, images or code commercially.
- Keep our own confidential information and trade secrets out of public AI services (see Section 4).
- For AI-generated code, respect licences and scan for security and licensing issues before use.

10 Working with generative AI

10.1 Text and writing

Great for drafting, summarising, rephrasing and brainstorming. Always edit for accuracy, tone and confidentiality, and make the final wording your own.

10.2 Code

Useful for boilerplate, explanations and debugging. Review and test all generated code, never paste secrets or proprietary code into public tools, and check dependencies and licences.

10.3 Images, audio and video

Check usage rights and avoid generating misleading media or content depicting real people without a clear, legitimate basis. Label synthetic media where transparency requires it.

> **PROMPT HYGIENE**
>
> Write clear prompts, but remember: **the prompt itself is data**. Don't include anything in a prompt that you wouldn't want the provider to see. Strip out names and identifiers where they aren't needed.

11 Human oversight and accountability

- For any decision that affects a person's rights, money, safety or opportunities, a **human must review and be able to override** the AI.
- Do not let an AI system take consequential actions automatically without appropriate checks.
- The person who uses the AI output owns the result — including any error. When you sign off, you are vouching for it.
- If AI output will drive an important decision, keep enough record to explain how the decision was reached.

12 Guidance for specific roles

The rules above apply to everyone. These notes highlight what matters most in common roles.

Managers & team leads

Set expectations, make sure your team knows the approved tools, watch for shadow AI, and model responsible use. You are accountable for how your team uses AI.

Developers & engineers

Review, test and scan AI-generated code. Keep secrets and proprietary code out of public tools. Constrain what AI assistants and agents can access and do.

Marketing & content

Fact-check and edit AI drafts, respect copyright and brand voice, avoid misleading claims, and disclose AI-generated media where required.

HR & people functions

Be especially careful with personal data and with any AI that touches hiring or evaluation — bias and privacy risks are high, and human review is essential.

Customer-facing & support

Verify AI-suggested answers before sending, protect customer data, and make clear when a customer is dealing with an automated system.

13 When something goes wrong

Mistakes and near-misses happen. Reporting quickly limits harm — and you will not be penalised for raising a genuine concern in good faith.

Report promptly via **[reporting channel / contact]** if you notice any of the following:

- Confidential or personal data was entered into an AI tool by mistake.
- AI produced harmful, biased, false or damaging output that was used or shared.
- An AI tool or account behaves oddly, or you suspect it has been manipulated.
- You encounter a suspected deepfake, AI phishing attempt or misuse of AI.
- You are unsure whether something you did was allowed — ask, don't hide it.

REMEMBER

Early, honest reporting is always better than a hidden problem that grows. When in doubt, report and ask.

14 Quick reference — do and don't

Always

- Use approved tools.
- Verify important output.
- Protect personal & confidential data.
- Stay accountable for the result.
- Disclose AI use when it matters.
- Keep a human in the loop.
- Report problems early.

Never

- Paste RED data into public AI.
- Trust AI facts or citations blindly.
- Use unapproved shadow AI.
- Share accounts or API keys.
- Let AI decide things about people alone.
- Pass off unchecked AI work as verified.
- Hide a mistake.

15 Employee acknowledgement

By signing below, I confirm that I have read and understood the Responsible AI Use Guidelines and agree to follow them when using AI systems for work.

Name	**[Full name]**
Role / department	**[Role — Department]**
Signature	
Date	**[DD Month 2026]**

A COMPANION DOCUMENT

These guidelines summarise responsible-use practices for everyday work. They sit under, and do not replace, the organisation's **AI Security Policy** and related policies. Where a specific rule or law is stricter, follow the stricter requirement, and ask the security team if anything is unclear.