AidGPT

GOVERNANCE GUIDANCE

AI Disclosure and Use

Guidance for managers, with a staff-notice template ready to adapt

MI-GUIDE-001, version 2.3 · External · August 2026 · Licensed under CC BY 4.0 · aidgpt.org

The one idea

Your staff are already using AI. The question is not whether to permit it, but whether you can see it.

Disclosure is the visibility control that lets managers govern AI use they cannot otherwise see. It turns an invisible practice into a managed one and can often be introduced without buying new software or waiting for a complete global policy.

The staff-notice template near the end is ready to adapt and route through your normal human-resources, data-protection and policy approval process.

Who this is for

A manager who is accountable for work that other people produce. A country director, a head of programmes, a research or analysis lead. You suspect AI is already in use in your team. You have no policy, or you have one nobody follows.

It assumes you cannot buy new tools this quarter, cannot wait for headquarters, and cannot credibly ban AI use outright.

It is not a legal document and it does not tell you what the law requires of you. It tells you what to do first, which is a different question and a more urgent one.

Why disclosure comes first

Many organisations work in the wrong order. Tools, then policy, then training, and disclosure last if at all.

That order fails because the first three depend on knowing what people are actually doing. Without disclosure you are managing a practice you cannot see. You find the problem when a donor or a peer finds it, which is the most expensive moment to find out.

There is a second reason, and for analysis teams it is the serious one. The risk is not only that staff use AI badly. It is that they produce work above their own capacity, cannot check it, and neither can anyone downstream. When an expert reads that work, finds an error and concludes it was machine-generated, the damage is not to the document. It is to the question of why the organisation should be commissioned at all.

An organisation whose product is analysis sells verified judgement. That is what disclosure protects.

The rule

State it in one sentence and do not soften it.

Staff may use approved AI tools for permitted work. Staff must disclose substantive AI assistance, in writing, to the person who reviews the work.

Substantive means AI use that materially shaped the research, analysis, structure, evidence, argument, wording or conclusions of an output, or created a material new record such as a meeting transcript. Routine spelling, grammar and predictive-text corrections do not need a note. If the tool made a substantive choice about the content, structure, evidence or argument, or created a material new record, it does.

Four things make the rule work.

- **It is permissive.** It gives staff a safe route to use approved tools for permitted work. It asks them to be visible about substantive assistance.
- **It is specific.** "I used AI" is not a disclosure. The test is whether a reviewer can tell what to check.
- **It applies to everyone.** Senior staff who use AI well disclose on the same terms as junior staff who use it badly. A rule that only points downwards reads as surveillance and gets ignored.
- **It is universal, which stops honesty being optional.** Optional disclosure rewards silence: the person who discloses carries a visibility cost that the person who conceals avoids. A common rule removes that advantage. It does not eliminate the social penalty attached to visible AI use, so managers must apply a common review standard and senior staff must disclose first. The evidence is set out in the accompanying essay, *Make AI Disclosure Boring*.

What a disclosure actually looks like

A common failure is a disclosure that discloses nothing. Below are three that work, for three documents many organisations produce. The pattern is the same each time: name the tool, name what it touched, name what you checked, take responsibility.

On a report or analysis

> In preparing this report I used Claude to summarise partner submissions and to draft the narrative sections. All figures were checked against the source data by me. All content was reviewed and verified by me and I take full responsibility for its accuracy and presentation.

On a terms of reference or a proposal

> AI assistance was used in drafting this document. Structure and wording were developed with Claude from an outline I wrote. The budget, the methodology and the personnel sections are my own and were not AI-generated. I take full responsibility for the content.

On an evaluation or a piece of research

> Claude was used, in an approved environment, to code and cluster de-identified open-ended survey responses and to draft the literature summary. I reviewed every code assignment against the approved de-identified source set, and every reference cited was checked against the original document. Interpretation and conclusions are mine.

And here is the same disclosure written badly, so people can see the difference:

- × "AI was used in the preparation of this document."
- × "This report was drafted with the assistance of artificial intelligence tools."
- × "Parts of this document may have been generated using AI."

None of those tells a reviewer where to look. That is the only job a disclosure has.

Three things people mean by disclosure

The word covers three different practices, and collapsing them causes trouble.

Workflow disclosure. The author tells the internal reviewer what AI did, what material it touched and what was checked. This is the rule in this note, and it is the minimum.

Methodological disclosure. Where AI materially affected the research, coding or analysis, the document itself explains the use, because readers need it to assess the method.

Public labelling. A publication tells its external audience that AI was involved. Whether that is needed depends on the type of publication, the nature of the contribution, contractual requirements and applicable rules.

These are related, but they are not identical. Every substantive use is visible to the internal reviewer. Not every spelling correction is declared to a donor. A methodologically significant use is not hidden behind a vague internal note. The purpose determines the disclosure.

What may go into an AI tool, and who decides

Disclosure governs visibility and accountability. Permission governs what the system may touch.

Clear, non-negotiable prohibitions may be necessary for particular data or uses. A permission gate should sit alongside them. What does not work is relying on a long, generic banned-material list without naming who classifies the material, approves the tool and decides the specific use. That leaves the decision at the wrong desk. An analyst deciding alone whether a dataset is safe is doing a risk assessment the organisation has not properly assigned or recorded.

The rule is a permission gate, not a list.

> **Public and routine internal material may be used only where the organisation's classification rules and applicable agreements permit it, and only in an approved tool. Anything beyond that requires two approvals: the tool, for that class of data, and the specific use.**

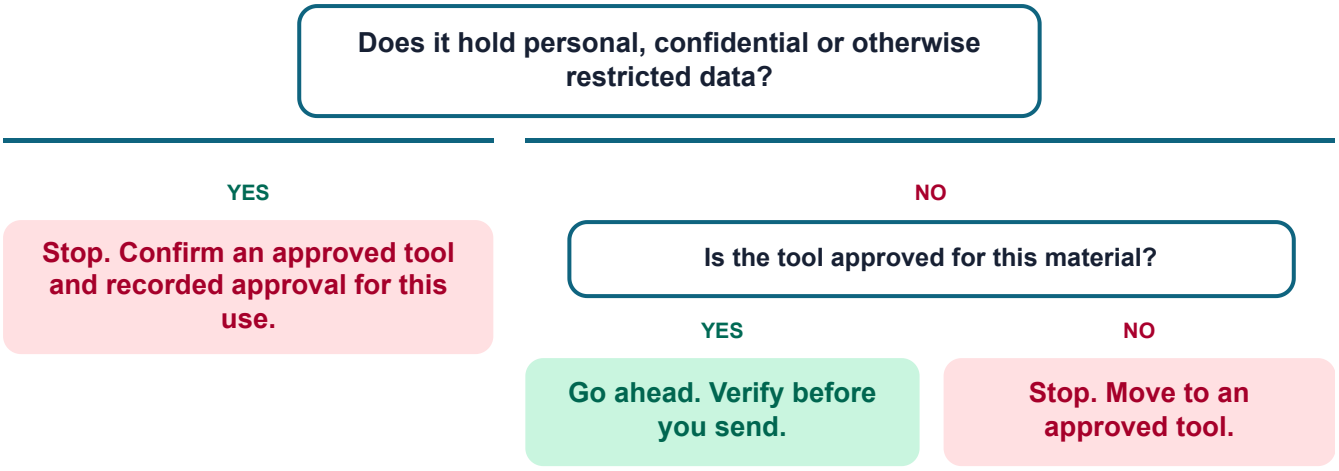
Those are two different decisions and usually two different people.

- **The tool is approved centrally.** Whether a given tool may hold a given class of data is a security and data-protection decision. It is written down and reassessed when the product, terms, data class or intended use materially changes. It belongs to the designated security or data-protection function, not to a line manager or analyst acting alone.

- **The use is approved by the designated accountable owner.** This is the role your organisation has made accountable for the dataset or processing purpose, with data-protection, information-security, safeguarding or legal advice where required. The role is not always the individual's line manager. It decides whether this particular use is necessary and proportionate and records the decision.

The analyst's job is to classify and ask. It is not to make the call.

Permission decision: check the material, the approved tool and the specific use before proceeding.



1. If the material includes personal, confidential or otherwise restricted data, stop. confirm an approved tool and recorded approval for this use.
2. If it does not, ask whether the tool is approved for this material. If yes, go ahead. verify before you send. If no, stop. move to an approved tool.

When unsure, pause and ask. Do not enter the material until you have an answer from the designated accountable role. An organisation where people ask is safer than one with a longer list.

The honest practical answer

In many country offices using cloud chat on a standard licence, the answer for personal data about affected people will usually be no.

Say that plainly, because a permission gate can read as though everything is negotiable. It is not. The difference is that the no is now a recorded decision by someone accountable, made against a classification and a named tool, rather than a rule an analyst reads off a poster and works around when it gets in the way.

The tier question is not free versus paid

This is a common procurement mistake, and it is expensive because it feels like a solution.

Managers buy everyone an individual paid subscription, assume the data question is now answered, and move on. It is not answered. The line that matters is not free against paid. It is consumer against commercial.

Take one of the tools used to prepare this note. At the time of publication, Claude Free, Pro and Max users control model-improvement use in their own privacy settings, although Anthropic says safety-flagged conversations may still be used for trust-and-safety work. Claude for Work Team and Enterprise accounts are governed by commercial terms under which Anthropic says inputs and outputs are not used for model

training by default, unless the customer explicitly opts in or submits material as feedback. So an office that buys twenty individual Pro subscriptions has bought twenty consumer accounts with settings individual users can change. It has not bought organisation-controlled governance.

Every vendor draws that line in its own place, and terms, features and product boundaries can change over time. The answer for one product on one date is not worth memorising. The questions are.

- Does the vendor use what we enter to train models, and is that default on or default off?
- Is that controlled by the organisation or by each individual user?
- Is it written into a contract we hold, or into consumer terms the vendor can vary?
- How long is our content retained, and does that change if the setting changes?
- If someone asks us to see or delete their personal data, can we actually do it inside the tool?

That last one is a useful way to test whether a tool is usable for personal data at all, and it is a question often skipped.

There is a second reason consumer accounts cause trouble. They may support Projects, uploaded documents and saved instructions, but they generally lack organisation-controlled governance such as central administration, audit logs, policy enforcement, organisational exports and retention controls. Staff are left to manage settings and records individually, with no reliable way for the organisation to review the set-up.

If AI use is already happening, providing organisation-approved tools is a control rather than a perk. Organisation-unapproved consumer accounts should not be used for work material, including public material, unless the organisation has explicitly authorised that use.

"Approved" does not necessarily mean an enterprise product has already been procured. Where no approved organisational tool exists, the designated accountable roles can issue a recorded, time-limited interim authorisation for a named service and account type. Keep it to defined low-risk tasks with genuinely public or synthetic material. Set the required privacy settings. Exclude connectors, file uploads, transcription and any personal, confidential, entrusted or otherwise restricted material. Name the owner and expiry date, and require disclosure and human verification.

If nobody has authority to accept even that limited risk, new AI use pauses until someone does. During this calibration period, give staff a confidential route to disclose existing use. Disclosure alone should not trigger automatic punishment.

AI you did not procure

Most of this note assumes someone chose to use an AI tool. A large part of the actual exposure in a typical office involves nobody choosing anything.

Productivity suites, video conferencing platforms and collaboration tools now ship with AI built in. Meeting transcription and summarisation. Email and document summarisation. Drafting assistance inside the word processor. Search across the organisation's own files. Software that quietly adds an AI function after an update.

Three things follow, and each one matters more than it first appears.

Your staff do not think they are using AI. Someone who would never paste a protection case note into a chatbot will happily let the meeting tool auto-summarise a case conference. They are not being reckless. In their mental model they have not used AI at all, so the disclosure rule does not occur to them and the permission gate is never approached.

A transcript is a new record you did not decide to create. A meeting that used to leave a set of handwritten notes now leaves a verbatim account of who said what, held somewhere and retained under settings nobody in the meeting chose. It may be accessible for legal, investigative, audit or data-access purposes, depending on the applicable regime. For protection, safeguarding, HR and investigation meetings, that is a material change in your risk position and it happened without a procurement decision.

The defaults are not yours to assume. Whether these features are on, who can turn them on, where the content goes and how long it is kept are tenant-level settings that differ by vendor, by licence and by version, and they change. Do not take anyone's word for what your defaults are, including this note's.

The first action is short and need not wait for a complete policy.

- Ask whoever administers your productivity suite and your video platform which AI features are currently enabled for your tenant, who can enable them, and what happens to the content.
- Decide which meeting types must never be auto-transcribed, and turn it off for those rather than asking people to remember. Instructions guide behaviour. Settings enforce boundaries.
- Say in the staff notice that AI built into the tools you already use counts as AI use.

Verification before anything leaves

Disclosure tells you what was done. Verification determines whether the output is ready to publish. Three requirements, in order.

- ✓ **A named person reads it.** Not skims, reads. If nobody will read the whole thing, nobody should send it. An AI can generate fifty pages in a minute. That does not create a human capacity to review fifty pages in a minute. This catches many avoidable failures and is often skipped.
- ✓ **Every factual claim is traced to a source.** Figures against the data, quotations against the record, references against the actual document. AI text is fluent whether or not it is true, so fluency tells you nothing. A model's confident tone is not a confidence rating.
- ✓ **The reviewer is competent to judge it.** If someone produces analysis in a field they do not know, and it is reviewed by someone who also does not know it, the review is decorative. Route it to someone who can judge it, or do not publish it.

Four-step workflow from permitted use to named responsibility.

The third requirement is the one organisations quietly fail. It is also the one that decides whether the first two were worth anything.

A disclosure is self-report, not an audit log. For routine, low-risk work, a short statement and ordinary editorial review are enough. Higher-risk work needs a record behind the statement: links to the source material, the checked dataset, a version history, the reviewer's name, a sign-off showing who verified which claims. Saying something happened is not evidence that it happened.

What managers do differently

Training staff to use AI without changing how you supervise them makes the problem harder to see, not smaller.

- **Read the disclosure first.** It tells you where to spend your attention.
- **Ask the follow-up questions.** "Where did this figure come from?", "what does this paragraph mean in your own words?" and "what did the tool leave out?" separate understood work from passed-through work. They are diagnostic questions, not a test of fluency: someone working in a second language may understand more than they can explain quickly. Use them to locate the review, then check the evidence.
- **Change what you check for.** The cost of producing text has collapsed. The cost of checking it has not. Time that went on drafting should now go on source checking, and workloads should say so rather than absorbing it quietly. If every minute saved in drafting becomes capacity for more output, the volume of AI-assisted work will exceed your ability to review it.
- **Answer the data question rather than deflecting it.** If you own a dataset, you are now the gate. Saying yes without thinking and refusing without reason are the same failure, and both push people back to doing it quietly.
- **Watch the three risks nobody warns you about.** One person hands over the judgement they are paid to make, and the output looks the same until something breaks. Another was already at capacity, and now uses AI to push past the limit that was protecting them. And the first-draft work an AI absorbs is often the work junior staff learn through, so today's team gets faster while tomorrow's organisation loses its pipeline. You will not find any of the three in an output review.

Managers also need to be fluent enough to supervise credibly. A reviewer who does not know what these tools can and cannot do will not know what to look for.

When someone does not disclose

The rule needs a consequence or it is advice.

Non-disclosure is not only a technical lapse. Once the organisation has clearly communicated that approved AI use is permitted provided it is disclosed, knowing concealment can misrepresent how the work was produced. Deliberate concealment after the rule is understood may be handled under the organisation's applicable conduct procedures, subject to local employment law, collective agreements and due process.

Explain this when you issue the rule. It is fairer to state the possible consequence and the process at the start than to improvise the first time a case arises.

What to do the first time it happens is a different question. In the first months after a new rule, an undisclosed output is usually evidence that the rule did not reach the person, not evidence of concealment. Treat the first cases as a test of your communication rather than of their conduct, find out what they understood the rule to be, and fix that. Reserve the conduct route for someone who knew and chose not to.

Three ways to kill the rule in its first month: reward disclosure by increasing someone's workload because a tool now helps them, humiliate the first person who writes an awkward disclosure, or make honesty the trigger for an investigation while silence stays invisible. What you do with the first ten disclosures will matter more than the wording of the policy.

Managers who go straight to conduct on day one teach the office that the safest thing is to say nothing, which is the outcome the whole rule exists to prevent.

The first thirty days

The first visibility steps need not wait for a complete policy or procurement programme.

This week. Put a disclosure on your own next substantive document, before you ask anyone else to. Name the accountable role that questions and use-approval requests go to. Establish an interim list of approved tools and permitted material, including any narrow, time-limited authorisation for public or synthetic material, even if the list is short and the answer is mostly no. Ask your systems administrator which AI features are already switched on in your productivity suite and video platform. Adapt the staff notice and route it through your normal human-resources, data-protection and policy approval process.

Within two weeks. Issue the notice once the interim tool list, permission route and named contact exist. Ask your data-protection officer or privacy lead whether an impact assessment covers AI use on your programme data. If none does, record that finding and ask what assessment is required before higher-risk use continues. Confirm which tools are approved for which classes of material and how changes in features or terms will trigger reassessment.

Within a month. Change one review process so the disclosure is read before the document. Ask two managers to try the follow-up questions and tell you what happened. Look at whether anyone's workload was rebalanced to reflect that checking now takes longer than drafting.

Then review it. Put a date in the diary six months out. The tools, the law and your own practice will all have moved.

Staff notice, ready to adapt

Replace the bracketed text, complete your normal internal approvals and then issue it. It is deliberately short.

AI use in our work: what is required of you

You may use approved AI tools for permitted work. You must disclose substantive AI assistance.

Disclose to your reviewer. When you submit a report, analysis or other substantive piece of work, include a short note for the internal reviewer saying what AI did and what you did. Name the tool, identify what it touched, say what you checked and take responsibility. Routine spelling, grammar and predictive-text corrections do not need a note. Whether a disclosure also appears in the final publication is a separate methodological, contractual or publication decision.

Built-in AI counts. Meeting transcription and summarisation, drafting assistance in documents and email, and AI search across our files are AI use. Apply the same data permissions to them. Disclose them when they substantively shape the work or create a material new record.

Pause before using sensitive material. Public or routine internal material may be used only where our classification rules and agreements permit it, and only in an approved tool. Personal, confidential, protection, safeguarding, beneficiary, raw-transcript, HR, investigation, security or entrusted material requires both a tool approved for that class of information and recorded approval for the specific use from [accountable role or process]. Do not enter it while either approval is unresolved.

Verify before you send. Read the whole output. Trace every factual claim, figure, quotation and reference to its source. You must be able to justify the reasoning and evidence, or route the work to someone competent to do so.

Answer for it. If you send it, you are the author. A tool helping does not change that.

Using approved AI tools for permitted work is not the problem. Hiding substantive AI assistance is. Knowing, deliberate concealment after the rule is understood may be handled under our applicable conduct procedures and local requirements.

Questions to [name and role]. Applies from [date].

How we hold ourselves to this

We publish our own position rather than only advising on other people's.

MarketImpact's published position says every published piece and client deliverable is reviewed, verified and edited by a named person with direct domain expertise. It covers research synthesis, drafting, analysis and software development. Judgement is not delegated. It states that work is risk-tiered before AI touches it, only enterprise tools with contractual no-training guarantees are used, highly sensitive material is processed locally on company hardware, beneficiary-level data never enters AI tools, and higher-risk outputs have a documented sign-off.

Every published analysis carries a disclosure stating what AI did and what the author did. This note carries one on the last page.

The full position is at marketimpact.org/how-we-use-ai. Our internal AI Use Policy is available to clients and partners on request.

What this note does not cover, and where to go instead

This is a governance note about visibility and accountability. It is deliberately narrow, and it leaves out several things that matter.

The law. This note does not tell you what data-protection law or the EU AI Act require of you. Data-protection questions should go to your data-protection officer or privacy lead. AI Act scope and obligations should go to your legal or compliance lead. The Act applies in phases, while implementation guidance and some deadlines continue to evolve. The answer turns on facts about your organisation this note cannot know. If you are a UN agency or an organisation with privileges and immunities, the framing is different again.

Data responsibility in humanitarian operations, an established field with mature frameworks that AI governance should build on rather than duplicate. The AI question sits inside it, not beside it.

Procurement, security assessment and vendor due diligence beyond the questions earlier in this note.

Anything about affected people's own use of AI, or AI-mediated services delivered to them, which raises a different and harder set of questions than staff use.

Four places to go next. Check the current version of each, because they are actively revised.

- **OCHA Data Responsibility Guidelines,** January 2025, developed by the OCHA Centre for Humanitarian Data. The operational baseline, and revised more recently than many people assume.
- **Handbook on Data Protection in Humanitarian Action,** ICRC, edited by Massimo Marelli, third edition, Cambridge University Press, 2024. Open access. Chapter 17 covers AI and machine learning broadly rather than providing product-specific guidance on today's generative-AI services.
- **Building a Responsible Humanitarian Approach: The ICRC's Policy on Artificial Intelligence,** November 2024. A humanitarian organisation setting out its own position, useful as a model as well as a source.
- **SAFE AI: Standards and Assurance Framework for Ethical AI in Humanitarian Action,** CDAC Network with the Alan Turing Institute and Humanitarian AI Advisory. Staged decision gates and three risk tiers, built for humanitarian action.

Using this document

If it is useful, use it. There is nothing to sign and nothing attached to it.

Version and review. Version 2.3, August 2026. Review date February 2027. The canonical publication URL is aidgpt.org/AidGPT-MI-GUIDE-001-AI-Disclosure-and-Use.pdf. Version 2.3 corrects the evidence claim about mandatory disclosure, strengthens the two-part permission gate, makes the interim-authorisation route explicit, aligns the staff notice with the guide, updates legal routing and product language, and adds accessible links and licensing instructions.

Our own disclosure. This note was drafted by Thomas Byrnes with AI assistance. Claude helped turn an outline into early sections and challenged the argument. OpenAI Codex audited the evidence and internal consistency, prepared the version 2.3 revisions, checked linked sources against publisher or official pages, and rebuilt the accessible publication files. An earlier draft carried a detailed account of the legal position. Parts of it were removed because they could not be traced to a source we had actually read. The structure, argument and conclusions are the author's, and he takes full responsibility for the content.