

HURIDOCs

Internal AI Governance
Policy and Framework



1. Purpose

This internal policy outlines how HURIDOCs staff, including team members, contractors, consultants, and interns, may use artificial intelligence (AI) tools in their day-to-day operations. It is designed to ensure that AI use is aligned with HURIDOCs' values, mission, and responsibility to human rights defenders and affected communities.

2. Scope

This policy applies to all HURIDOCs staff, including full-time employees, part-time staff, contractors, consultants, and interns. It covers any AI tool used to perform:

- Administrative tasks
- Communications (internal and external)
- Project management
- Software development
- Partner engagement and support
- Donor engagement and fundraising

3. Definitions

- **AI tools:** Any system or service that uses AI algorithms, such as machine learning or large language models to generate, interpret, or manipulate data or content.
- **Generative AI:** Tools that generate new content such as text, code, images, audio or video (e.g., ChatGPT, GitHub Copilot, Microsoft Copilot, Google Gemini, Claude, DALL-E, Midjourney).
- **Sensitive data:** Any information related to identifiable individuals, vulnerable communities, internal operations, or partner material. Partners get to define and state limits on what they consider sensitive when it comes to their data.



4. Principles

AI use at HURIDOCs should reflect the following principles:

Human Oversight: AI should augment, and not replace, human judgment. All AI-generated outputs must be critically reviewed before internal or external use.

Transparency: When AI-generated content is shared externally (with partners, donors, or the public), it must be clearly labelled and disclosed as such. When AI tools were used to augment (beyond spelling or grammatical corrections) or generate content for internal use, staff should disclose the tool and to which extent it was used. AI use disclosures should identify the AI tool used, clarify what the tool was used for, affirm that a human reviewed the output before sharing, and acknowledge that the activity is in accordance with HURIDOCs' AI Governance Policy and Framework. This promotes transparency, reinforces human oversight, and ensures that no sensitive or identifiable information was processed without proper safeguards.

Consent: Where documentation or communication involves identifiable individuals, consent norms must be respected. Consent needs to be secured from everyone involved and implicated.

Security and Privacy: AI tools must not be used to process sensitive or partner-related data unless reviewed and approved. When using translation tools, anonymisation techniques are required at all times, and any identifiers should be masked, pseudonymised, generalised, swapped, perturbed (altered or noise added), or synthetically

generated. Tools must comply with HURIDOCs' broader digital security protocols.

Equity and Access: Consideration should be given to accessibility affordances and constraints.

Environmental Responsibility: Staff are encouraged to use AI tools responsibly and efficiently, in recognition of their environmental footprint. Staff should be mindful that the carbon footprint of an AI query is at least 10 times higher than a traditional search query, while complex and generative AI queries are considered to have the biggest impact on the environment.

Staff are encouraged to minimise unnecessary AI usage especially for tasks that can be done without computational assistance. When using AI/ML tools, prioritise smaller, less resource-intensive queries, and avoid repeated or speculative prompts that strain server load. Choose tools that offer environmental transparency or green computing commitments. Where possible, use shared or centralised instances rather than individual accounts to optimise energy use. Teams should consider scheduling heavier AI tasks during off-peak hours and explore low-energy alternatives for recurring tasks, such as rule-based automations or offline tools. The organisation will continue to keep abreast with research and interventions around harm responsible AI use vis-à-vis its impact on the environment.

5. Acceptable Use

Staff may use approved AI tools to:

- Draft emails, documentation, or reports (with review)
- Translate or summarise non-sensitive content
- Generate ideas or outlines for content or campaigns
- Support programming and coding workflows
- Transcribe or record meetings (with consent)

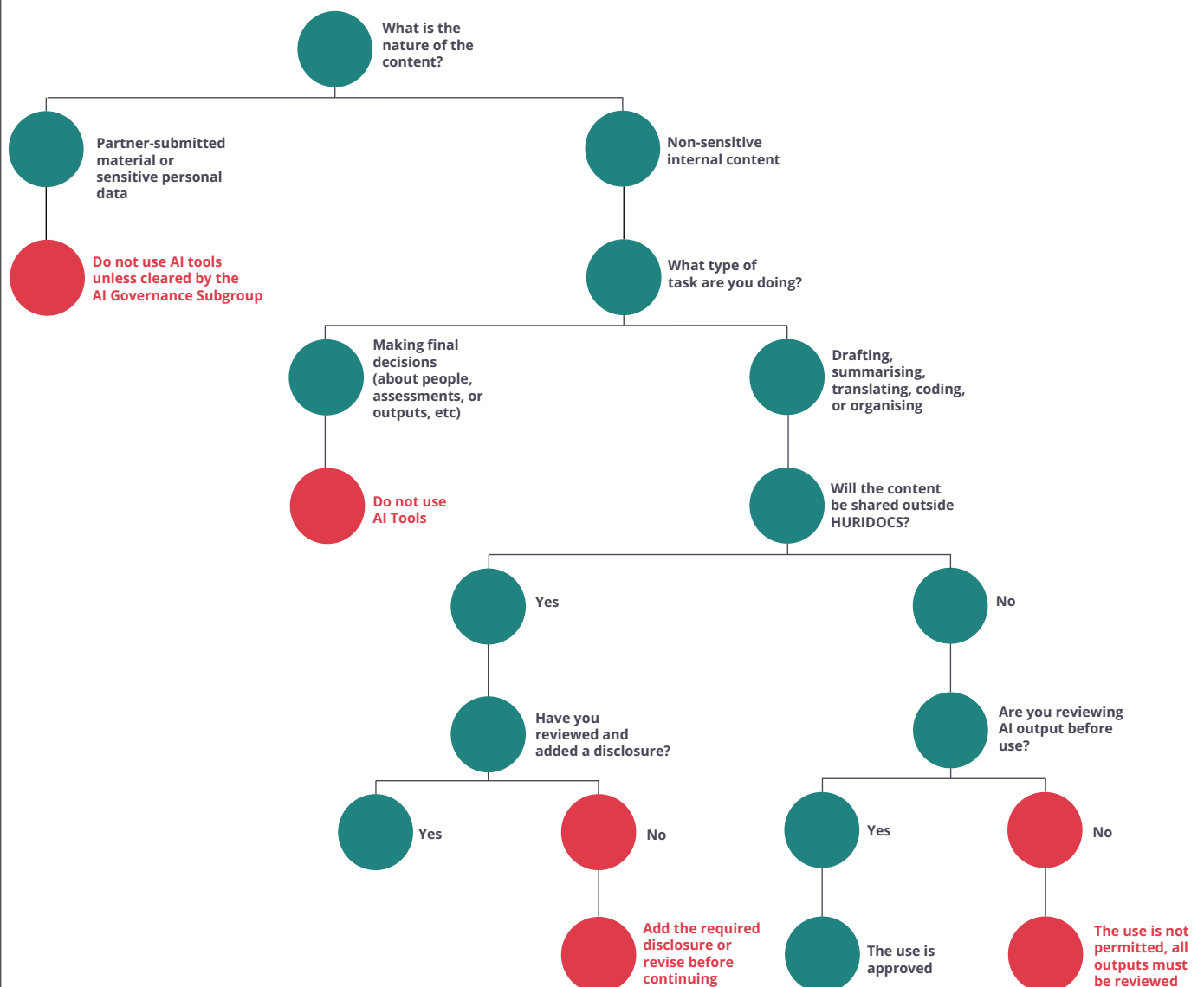
Staff should be transparent and disclose the type of AI tools used and to which extent it was used.

6. Prohibited Use

Staff must not use AI tools to:

- Process, store, or transmit sensitive or partner-submitted data unless formal consent/approval is obtained
- Make final decisions about people such as hiring, performance evaluation, personal security assessments, or termination of employment
- Generate reports, statements, media or content shared externally without review
- Represent organisational views, values, or commitments without human oversight

Responsible AI Use Decision Tree





7. Tool Selection and Approval

All AI tools used for work must be listed in an internal registry.

New tools must be evaluated by the Development and Communications Team (DevComms Team) with the AI Oversight Subgroup using the AI Tool Evaluation Template.

Selection should favour tools with strong data privacy, disclosure features, and reasonable environmental impacts.

8. Cost and Access

HURIDOCS, where possible, will secure and manage institutional licenses or subscriptions for selected and approved tools. User accounts will then be given to relevant team members.

Staff should not pay out of pocket for work-related AI use, unless pre-approved. If a new tool is deemed necessary for performing work-related tasks, this should be discussed with the DevComms Team with the AI Governance Subgroup.

Centralising procurement helps monitor usage, cost-efficiency, and compliance. Where staff use stipends or other personal funds for work-related AI tools, they must record the tool in a shared AI Tools Register for visibility and review. Tools used for recurring work, or any use involving

sensitive data, should be prioritised for institutional procurement and centrally managed accounts whenever feasible.

Tools purchased personally used for recurring work, or any use involving sensitive data, should be prioritised for institutional procurement and centrally managed accounts whenever feasible.

9. Training and Knowledge Sharing

HURIDOCS will support regular check-ins, learning exchanges, and onboarding sessions to surface AI-related concerns and practices.

Each team is encouraged to engage in peer-to-peer learning where they discuss tool use, prompt engineering, and general risks and insights in using these tools.

10. Team-Level Guidance

Team-specific guides provide differentiated guidance:

- Tech and Product Teams: coding, testing, debugging
- Programmes Team: configuration support, translation, synthesis
- Management & DevComms: strategic communications, administration, content generation, translation



II. Disclosure

AI use should be disclosed when it meaningfully shapes the output being shared, particularly when that output is visible beyond the individual or team, or when it contributes to decisions, narratives, or documentation with lasting or external impact.

Disclosures should include:

- What AI tool was used
- What the tool was used for
- Confirmation that a human reviewed and finalised the output
- A note that the activity complies with HURIDOCs' Internal AI Governance Policy and Framework

This reinforces human oversight, promotes transparency, and ensures no sensitive or identifiable information was processed without proper safeguards.

Below is an example of a disclosure statement:

"This document was [function] with the support of [tool] and finalised by a HURIDOCs team member. AI assistance was used in accordance with HURIDOCs' internal AI Use Policy and Framework."

We recognise that not all AI uses warrant formal disclosure. For example, lightweight or assistive uses like rephrasing an email draft, fixing a line of code, or generating placeholder text may not need to be flagged. However, if AI-generated content is being reused, shared outside the team, or relied on for significant decisions, disclosure is encouraged.

For code in particular: quick syntax help or suggestions may not need to be acknowledged. But if AI-assisted code is integrated into a shared codebase, published externally, or used in production environments, disclosure helps maintain trust and clarify authorship.

A brief note such as *"This script was generated with the help of GitHub Copilot and reviewed by [name]"* is sufficient.

The aim is not to be exhaustive, but to foster habits of intentionality and care. When in doubt, reflect: Did AI meaningfully shape the final output? Is this something others will rely on? Could naming its use promote accountability or mitigate misunderstanding? If yes, a quick disclosure helps.

12. Infrastructure and Security Integration

All AI tool usage must comply with HURIDOCs' digital security policies.

Tools embedded in infrastructure (e.g. Google Workspace, Zoom) must be audited for AI features.

HURIDOCs will continue to assess whether self-hosted or open-source models are preferable.

13. Governance and Review

The DevComms Team will coordinate the first draft and revisions of this policy.

A cross-team AI Oversight Subgroup will provide tool reviews, track usage patterns, and lead updates. The AI Oversight Subgroup will be coordinated by an AI Oversight Coordinator.

The policy will be reviewed every 6 months, or sooner as needed.

14. Living Framework

This document is versioned and meant to evolve. It is not a rigid rulebook but a starting point for accountable, responsible, and just use of AI in human rights work.

For questions or contributions to the HURIDOCs Internal AI Policy and Framework, please contact us through our Senior Documentalist and AI Governance Subgroup Coordinator, Bono Olgado at bono@huridocs.org.

As an exercise, this policy was initially drafted with the use of ChatGPT primarily to frame/structure the policy by identifying key areas for discussion and shaping it in the genre of a policy document. Input included findings from previous team surveys and discussions. The output was revised through several prompt exchanges and then reviewed, annotated, and edited closely by members of the HURIDOCs team.



HURIDOCs

Human Rights Information and
Documentation Systems

HURIDOCs INTERNAL AI GOVERNANCE
POLICY AND FRAMEWORK

huridocs.org