Globethics

Responsible AI Use Policy

Approved by the Board of Foundation on 23 October 2026

1. Purpose

Globethics has been operating at the intersection of ethics and practice since the beginning, developing solutions to ethical dilemmas and moving beyond compliance towards impactful, values-driven governance and action across sectors. As artificial intelligence becomes a routine part of how society works, researches, and communicates, this policy sets out how Globethics uses it.

The Responsible AI Use Policy aims to give the clarity and confidence to experiment, innovate, and leverage the potential of AI tools, while protecting the people we serve, the partners we work with, the donors who trusted us, and the reputation we have built.

This policy reaffirms Globethics as a place where staff feel free to ask questions, share what they are learning, and try new approaches. This policy also ensures that our practices comply with applicable law in Switzerland, including the New Federal Act on Data Protection (nFADP), at the regional level including the EU Artificial Intelligence Act (EU AI Act) and the General Data Protection Regulation (GDPR), and at the global level. It is also informed by global standards and principles on the ethics of AI, including the OECD AI Principles and the UNESCO Recommendation on the Ethics of Artificial Intelligence.

2. Scope

This policy applies to everyone who uses AI tools for any Globethics-related work, including:

- Staff (full-time, part-time, and remote)
- Mandate holders and consultants
- Board members and advisors
- Contractors and service providers
- Fellows and programme participants when working on Globethics activities
- Partners, where AI tools are used in the context of joint work

It covers AI tools you access yourself, not only systems deployed centrally by Globethics. If you are using any AI tool, whether free or paid, as part of your Globethics work, this policy applies.

It covers all types of work: internal operations, research, project implementation, communications, fundraising, training, external publications, among others.

3. Definitions

To support this policy use consistently, here is how we use key terms throughout the document.

Term	**What it means**
AI tool	Any software that uses artificial intelligence techniques to generate, summarise, translate, transcribe, classify, or otherwise process content.
AI-assisted use	Using an AI tool to support or improve your own work. For example, to check grammar, generate a first draft that you then substantially edit, or translate a document.

AI-generated content	Content where AI played a substantial role in creating the output. For example, a report section, image, or presentation slide produced primarily by an AI tool, with limited human editing.
Approved tool	An AI tool that has been reviewed and added to the Globethics approved tool register, maintained by the Administration and Finance Officer.
Artificial Intelligence (AI)	A machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.
Deep fake	AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.
Substantive AI use	AI played a meaningful role in shaping the content or decision, beyond simple formatting, grammar correction, brainstorming or translation.
High-risk use	Uses of AI that could significantly affect people's rights, wellbeing, or opportunities, and/or the rights, opportunities, or reputation of organisations.
Prohibited use	Uses that are not permitted under this policy or applicable law, regardless of context.
Public AI tool	Any AI tool accessed via a consumer-facing interface whose terms of service do not guarantee that user inputs are kept confidential, excluded from model training, or processed solely for the purpose of delivering the requested output. This includes free and paid consumer tiers of tools where the provider retains the right to use, review, or train on user-submitted content.

4. Encouraged Uses

We encourage staff to explore how AI tools can make work more efficient, creative, and impactful. We use AI in internal processes to enhance, not replace, human expertise. AI should also be used with consideration for its environmental and resource impacts, favouring purposeful and proportionate use. Encouraged uses include, but are not limited to, the following:

- Drafting and editing documents, emails, and presentations
- Translating content between languages
- Researching topics and gathering information
- Analysing information
- Summarising documents, books, reports or research materials
- Transcribing audio or video recordings
- Taking meeting notes. When external parties are joining a meeting, we must ask for their permission.
- Generating ideas, outlines, or structures for projects

- Editing videos and creating subtitles
- Proofreading, formatting, and grammar checking

For all uses, the core principle is human oversight: AI is a tool that supports your judgement, not a replacement for it.

AI-generated outputs work best as a starting point, not a final version. In the case of documents, your own perspective, knowledge of Globethics' circumstances, and judgement are what make the work useful to others, your personal edits make a big difference. You are responsible for the content you produce, regardless of how it was generated.

5. High-Risk Use Cases

Some uses of AI carry more significant risks than others. They could affect people and organisations' rights or opportunities, or produce outputs that reach a broad external audience.

5.1 High-Risk under the EU AI Act (Annex III)

The following are considered high-risk use cases under EU AI Act Annex III, several of which are relevant to Globethics' work. For these use cases, a prior risk assessment and approval is required before proceeding:

High-risk use area	Examples in our context	Required before use
Recruitment and selection, performance evaluation, promotion, and dismissal	Screening candidate applicants, shortlisting candidates for positions, evaluating eligibility for programmes	Risk assessment + AI Ethics Manager + Executive Director approvals
Education evaluation and admissions	AI-generated assessments, automated progress tracking, personalised learning recommendations that affect outcomes	Risk assessment + AI Ethics Manager approval
Decisions affecting rights or wellbeing	Any use of AI that could influence decisions about individuals' access to resources, opportunities, or support	Risk assessment + AI Ethics Manager + Executive Director approvals

5.2 High-risk uses under this policy

These were identified through staff consultation as uses that carry heightened risk in Globethics' specific operational context.

- AI used to generate content that will be published under Globethics's name in formal contexts including donor reports, academic or research publications, official communications to UN bodies or intergovernmental partners, and grant proposals. The reputational and credibility stakes in these contexts are high.
- AI used in the design or delivery of training materials where the content is presented as expert or authoritative. Given that Globethics's learning programmes reach public officials, civil

society organisations, and emerging leaders globally, errors or biases in AI-generated content could have downstream effects at scale.

- AI tools integrated into organisational workflows that process personal data, including tools adopted for project management, communications, or learning platforms where data about staff, fellows, donors or partners is involved.

For these use cases, a prior risk assessment is required and sign-off from the AI Ethics Manager before proceeding.

6. Prohibited Uses

The following uses are not permitted under this policy. They are prohibited either because they pose unacceptable risks to people, because they conflict with our values, or because they are legally prohibited under Swiss law, EU AI Act, or other applicable international legal frameworks.

6.1 Prohibited under the EU AI Act (Article 5)

These prohibitions apply to all organisations operating under EU law and cannot be waived:

- Using AI to manipulate people through subliminal or deceptive techniques that could harm them
- Using AI to exploit the vulnerabilities of individuals or groups based on age, disability, or social circumstances
- Using AI for social scoring, evaluating or classifying people based on behaviour or personal characteristics in ways that lead to detrimental or unjustified treatment
- Using AI systems for real-time remote biometric identification in publicly accessible spaces
- Using AI to infer political opinions, religious beliefs, sexual orientation, or other sensitive attributes from biometric or behavioural data
- Using AI for emotion recognition in the workplace or educational institutions

6.2 Prohibited under this policy

These uses are prohibited by Globethics regardless of legal requirements, because they conflict with our values or pose significant risks to trust, privacy, or integrity:

- Creating deep fakes and AI-generated images, videos, or audio featuring real, identifiable people
- Using an AI-generated voice to represent Globethics in public communications without disclosure and approval from the Communications Manager and the Executive Director
- Presenting AI-generated content as entirely your own original work in contexts where attribution is expected, including but not limited to academic or research publications
- Using AI to produce content that is misleading, discriminatory, or in violation of our Code of Ethics and Code of Conduct.

If you are unsure whether something falls under a prohibited use, ask before proceeding.

Globethics

7. Risk Assessment

A risk assessment is a structured exercise to help you identify and manage potential problems before using an AI tool in a high-risk context (see section 5 of this Policy). It does not need to be long or complicated.

A Globethics AI risk assessment should address at least the following questions:

1. Who could be affected by this use of AI, directly or indirectly?
2. Could the AI tool produce biased or unfair results? How would we know?
3. What happens if the tool makes an error? What are the consequences?
4. Do the people affected have a way to ask questions, seek human review, or challenge the outcome?
5. What data will be entered into the tool, and is it appropriate to do so?
6. Has the tool been approved for use at Globethics, or reviewed for data protection compliance?
7. Who is responsible for oversight?
8. How will the use be monitored over time?

A template for the risk assessment is available here. Completed assessments should be shared with the AI Ethics Manager and kept in a designated folder.

8. Data Protection and Confidentiality

When you type information into a public AI tool, that information may be stored, processed outside Switzerland or the EU, or used to train future versions of the model. For this reason, the following types of information must never be entered into a public AI tool:

- Personal data of any individual (contact details, identification numbers, or any other sensitive data that could identify a person)
- Confidential documents or internal communications marked as restricted
- Donor or partner data, including financial information, contact details, or project terms
- Grant agreements, contracts, or legally privileged material
- Internal financial or budget information
- Meeting notes or records involving external participants who have not consented to AI processing
- Staff personal information
- Any content shared in confidence by external parties

If you are unsure whether the information you are working with falls into one of these categories, treat it as confidential and do not enter it into a public AI tool. You can always ask the AI Ethics Manager.

9. Governance and Roles

Responsible AI use at Globethics is a shared responsibility. The structure below reflects everyone's roles:

Role	Responsibilities
Executive Director	Final authority on high-risk AI use decisions. Approves AI use in recruitment, selection, and legal matters where required. Receives escalated incidents.
AI Ethics Manager	Owns and updates this policy. Reviews and approves requests for high-risk AI uses. Reviews new AI tools proposed for organisational adoption. Receives incident reports and escalates to the Executive Director where necessary. Acts as the first point of contact for any questions about this policy.
Administration and Finance Officer	Maintains the Globethics approved AI tool register. Flags any new AI tool purchases or subscriptions to the AI Ethics Manager before adoption. Supports the technical review of tools where relevant.
Portfolio Leads and units or programme Managers	Approve AI-assisted workflows within their area of work. Support their teams in understanding and applying this policy.
All staff, mandate holders, board members, fellows, and contractors	Use AI tools responsibly and in line with this policy. Report any incidents, concerns, or unexpected AI-related risks to the AI Ethics Manager. Seek approval before using a new AI tool not already on the Globethics approved AI tool register. Ask questions freely.

10. AI Tool Adoption

As AI tools continue to evolve rapidly, it is important that Globethics has a process for adopting AI tools. The goal is to make sure new tools apart from those licensed by the Head Office are safe, appropriate, and do not create unexpected risks.

10.1 Approved AI tool register

The Administration and Finance Officer maintains a register of AI tools approved for use at Globethics. The register is a spreadsheet, accessible to all staff, with one row per tool. It is the first place to check before using any AI tool for Globethics-related work.

For each tool, the register includes:

- Tool name and a one-line description of what it does
- Approved scope of use, whether the tool is approved for use with internal content, or only with non-sensitive content
- Data processing status, whether a Data Processing Agreement or enterprise terms are in place, or whether the tool operates under standard consumer terms
- Conditions of use, any specific restrictions, for example "do not enter personal data" or "approved for translation tasks only"
- Date last reviewed

The AI Ethics Manager reviews the submission, in coordination with the Administration and Finance Officer, and notifies the requesting staff member of the outcome. The burden of providing initial information rests with the staff member proposing the tool, not with the AI Ethics Manager or IT.

10.2 Requesting a new tool

If you would like to use an AI tool that is not already on the approved AI tool register, the process is the following:

- Contact the AI Ethics Manager, keeping in copy the Administration and Finance Officer, informing that you have added a new request for review to the AI Tool Register with the name of the tool, accessible link, whether a free or paid tier would be used, what you would use it for, and confirmed that: (i) you have reviewed the tool's privacy policy and terms of service, and (ii) you have searched on the Internet for recent news about the AI tool looking for incidents, and flag any relevant findings.
- The AI Ethics Manager, in coordination with the Administration and Finance Officer, will review the request and respond within a reasonable timeframe.
- If approved, the tool shall be available for use. If not approved, the AI Ethics Manager will explain why and suggest alternatives where possible.

Staff are encouraged to share tools they have found useful. The review process is collaborative.

11. Transparency and Disclosure

Being transparent about how we use AI is part of practising what we preach as an ethics organisation. At the same time, disclosure requirements should be practical and proportionate.

Here is how disclosure applies to different types of output:

Output type	Disclosure requirement
Academic or research publications	Always required. Disclose in the methods section or acknowledgements.
Reports to donors	Required if AI played a substantive role. Add a brief note.
Grant proposals	Required if AI played a substantive role. Add a brief note.
Training materials and learning content	Required if AI played a substantive role, given the potential impact on learners.
Recruitment or evaluation materials	Always required. AI use in these contexts carries heightened accountability.
External newsletters and website content	Required if AI played a substantive role.
Internal documents and reports	Required if AI played a substantive role. A note in the document or accompanying email suffices.
Social media posts	Judgement call. Not required for routine posts; recommended if AI generated most of the content.
Emails	Not required for AI-assisted drafting or grammar checking.

Translated documents	Not required unless the translation is published in a formal context (e.g., an official report).
Grammar checking, formatting, routine editing	No disclosure required.

Disclosure does not need to be elaborate. A sentence such as 'This document was drafted with AI assistance and reviewed by [name, role]' is sufficient in most cases.

12. Human Oversight

AI is a tool that supports human judgement, not a substitute for it. Regardless of how an output was generated, the person who submits, publishes, or acts on that output is responsible for it. This means:

- All AI-generated outputs must be reviewed by a human before they are used at work, shared externally, or published.
- AI should be used as an input and a starting point, not as a sole or final source.
- Where AI assists in a decision that affects someone else, that person has the right to know, to ask questions, and to request a human review of the outcome.
- Portfolio leads and units or programme managers are responsible for ensuring human review of AI-assisted workflows within their teams.

This principle exists because we believe that human judgement, ethical reasoning, and accountability are irreplaceable and because the people affected by our work deserve no less.

13. Incident Reporting

If something goes wrong or if you notice something that might go wrong, we must be aware of it and address it promptly.

An AI-related incident or near-miss might include:

- Confidential information accidentally entered into a public AI tool
- An AI-generated output that was shared externally without review and contained errors or misinformation
- An AI tool producing biased, discriminatory, or harmful outputs
- A data breach or privacy concern related to AI tool use
- Use of an AI tool in a way that is inconsistent with this policy
- An observed gap in how your team or other staff is using AI tools. For example, noticing colleagues using an unapproved tool, or concern that awareness about data confidentiality is unclear.

How to report:

- Complete the AI Incident Report Form available here, as soon as you become aware of the incident. The form is available online and will gather the information needed to assess and respond to the incident. The form guides you through describing what happened, when, who

was affected, what data was involved, what severity level the incident represents, and any immediate actions you have already taken.

- For critical incidents (those involving immediate risk to people's rights, wellbeing, or opportunities, or significant legal or reputational impact), contact the AI Ethics Manager directly by email or phone immediately in addition to submitting the form. Do not wait to complete the full form if the situation is urgent.

You may submit the form anonymously if you prefer, though identified reporting allows us to follow up if we need clarification

- The AI Ethics Manager will assess the situation and escalate to the Executive Director if necessary.
- Where required by law, Globethics will notify the relevant authorities within the required timeframe.

Incidents and near-misses will be used to improve our guidance and practices.

14. Training and Awareness

AI tools are changing fast. What is best practice today may need to be updated in the next months. Globethics is committed to keeping staff informed and supported as the landscape evolves.

- An introduction to this policy will be provided to all staff, mandate holders, and relevant contractors when it comes into effect.
- Regular updates, AI Ethics coffees, and learning sessions will be organised by the AI Ethics Manager to share new developments, tools, and guidance.
- Globethics Academy will incorporate training on AI periodically.
- Staff are encouraged to share tools they are using, questions they have, and experiments they are running. An open channel for AI-related questions and discussions will be maintained.
- Anyone with questions about this policy at any time is welcome to contact the AI Ethics Manager directly.

We see AI literacy as a core competency for the organisation going forward, and we are committed to building it together.

15. Policy Review

This policy will be reviewed annually by the AI Ethics Manager, in consultation with the Executive Director and the Global Management Team, and when required, is submitted for board approval.

An out-of-cycle review may be triggered by:

- Significant changes in the AI tools Globethics uses
- New or updated regulation, including updates to the Swiss regulation, EU AI Act, or new regulatory frameworks.

- A significant incident or near-miss
- Substantial feedback from staff

Staff feedback on this policy is always welcome and will be considered in any revision.

- Prepared by the manager of the Ethics of AI and emerging technologies portfolio, in consultation with the team
- Reviewed by the Executive director
- Trial version between 10 August and 14 September 2026, for further feedback from the team
- Updated version presented to the Board for validation on 23 October 2026