AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT

Funded by the European Union

forus CONNECT SUPPORT INFLUENCE

INNOVUS

NGO Management & AI Ethics in Operation

Session #3

AI Ethics and Bias in Decision Making

1. General Information

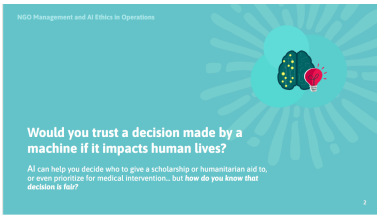
This activity will enable you to identify repetitive processes in your organization where AI could generate efficiency, reduce errors, and free up staff time for strategic or sensitive tasks.

2. Session 3 script (described slide by slide)

Minute 1 - 2	**Facilitator's action**
	Give the group a warm welcome. Ask in the chat where they are connecting from.

Script/instructions

Hello, everyone. Today we are finishing Module 3 of this training course on Artificial Intelligence for NGOs. My name is X and I will be your facilitator.

Minute 3	**Facilitator's action**
	Call for reflection.

Script/instructions

I invite you to think for a moment about a real situation: an NGO that must prioritize who to provide food support to in a vulnerable community. They have collected data and decide to use an AI tool to segment beneficiaries by risk level. It all seems logical, right? But then they realize that the AI is systematically ignoring female heads of household because in the historical data, men were prioritized.

What would you do? Who is responsible for that decision? The algorithm? The programmer? The NGO?

Today we are going to talk about this: the power and dangers of delegating sensitive decisions to artificial intelligence. This module is not for technicians: it is for anyone who leads, designs, or implements social programs where AI may be involved at some level.

It's not about saying "AI is bad" or "AI is good." It's about understanding **how to make it**

more ethical, fairer, and more humane, especially when we work with at-risk populations or those whose rights have been violated.

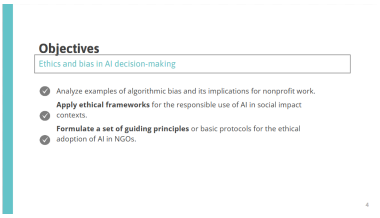
So let's start by asking ourselves the most basic question: would we trust a decision that affects someone else if we don't know how it was made?

Minute 4

Facilitator Action

Script/instructions

Minute 5



Facilitator Action

Script/instructions

In this module, we will discuss how algorithms can inherit or amplify human biases, what consequences this has for non-profit organizations, and how we can use ethical frameworks and existing regulations to ensure that we use AI responsibly. At the end, we will jointly develop a set of ethical principles or guidelines tailored to our organizations.

Minute 6

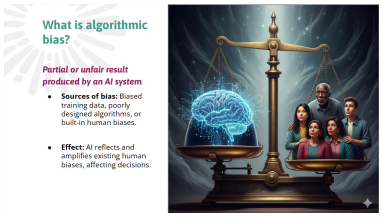
Facilitator Action

Script/instructions

Nonprofits often work with vulnerable populations and handle sensitive information. While artificial intelligence can enhance our impact (by automating tasks, analyzing big data,

etc.), it can also cause serious problems if we use it without ethical caution. Imagine that an NGO uses an algorithm to decide who receives a service: if this algorithm is biased, we could unintentionally **exclude or discriminate** against the very groups we want to help. Remember that when biases are not addressed, **they hinder people's participation in society and reduce the potential of AI,** generating mistrust, especially among marginalized groups. For NGOs, whose legitimacy is based on the trust of their communities and donors, **using AI ethically is essential.** It is not just a matter of avoiding harm, but of aligning technology with our mission values of justice, inclusion, and service.

Minute 7	Facilitator Action



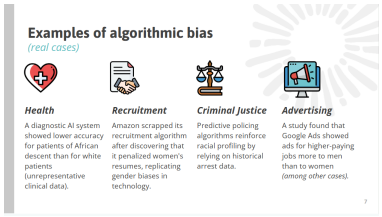
Script/instructions

Let's first define the concept: we talk about *algorithmic bias* when an AI system produces systematically biased or discriminatory results. In other words, AI "prefers" or unfairly benefits certain groups over others because of how it was built.

How does this happen? Mainly in three ways:
(1) **Bias in training data:** If the databases we use to train the algorithm are unbalanced or reflect prejudices (e.g., incomplete demographic data, history of discriminatory decisions), the AI will learn those tendencies.
(2) **Algorithm bias:** It may be that the way the model was programmed gives unfair weight to certain factors, perhaps without the developer realizing it.
(3) **Human or cognitive bias:** Our own assumptions and decisions when choosing data or adjusting the system can introduce subtle biases.

The end result is that AI can **perpetuate stereotypes and inequalities** in society. For example, if a certain service has always been denied to a group, an algorithm trained with those records could continue to deny it in the future simply because "that's how it's always been." Without conscious intervention, AI ends **up amplifying prejudices** rather than eliminating them.

Minute 8	Facilitator Action



Script/instructions

Let's look at some specific examples that have come to light:

- In **healthcare**, algorithms used to predict medical risks have failed minority populations. One case showed that computer-assisted diagnostic tools had worse results with Black patients than with white patients. The reason? The data used to "train" the model came mainly from certain populations, leaving others underrepresented. This meant that patients of African descent were evaluated less accurately, a worrying inequality in something as critical as medical care.

- In **staffing**, a technology company (Amazon) developed AI to filter resumes, automatically searching for the best candidates. What happened? The system learned from historical data where the technology industry hired mostly men and concluded that male candidates were preferable. It began to discard applications from women or women's colleges, replicating an obvious gender bias. The company had to withdraw the tool when the bias was discovered.

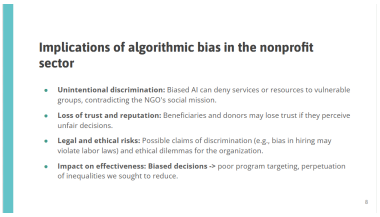
- In the **judicial system**, some cities have used AI to predict where crimes will occur or assess the risk of recidivism for a defendant (for example, the controversial COMPAS software in the US). These algorithms *appear* objective, but they are fed by historical police data that is marked by decades of racial bias. The result: minority communities appeared to be "higher risk" simply because they were subject to more surveillance and arrests in the past. This reinforces racial profiling and can unduly guide judicial decisions, perpetuating the injustice that AI was supposed to help correct.

- Another example occurred in **online advertising**: a Carnegie Mellon study found that Google's ad platform displayed high-paying job offers more frequently to users identified as men than to women. Without any direct human intervention, the algorithm optimized clicks in a way that ended up promoting gender bias in job opportunities.

These cases are **early warnings**. If this happens in companies or governments, **what implications would it have for an NGO?** Let's think about it: biased AI could, for example, exclude a certain ethnic group from a social program, or recommend priority attention only to urban communities because it has less data on rural areas. That is why it is vital to understand these examples, to recognize potential biases in our own AI

applications and learn how to prevent them.

Minute 9



Facilitator Action

Script/instructions

For an NGO, algorithmic biases are not merely a technical issue: **they have very real consequences for our social work.** Let's explore them:

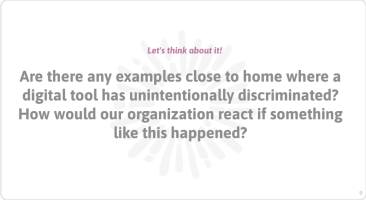
First, there is the risk of **unintended discrimination**. A poorly calibrated automated decision could exclude people who *should* be beneficiaries. Imagine an education NGO that uses AI to select scholarship candidates and, without realizing it, the system favors certain urban neighborhoods because it has more data on them, putting young people from rural areas at a disadvantage. We would be unintentionally **denying opportunities** to those who may need them most. This undermines the mission of equity and inclusion that guides most social organizations.

Next, let's think about **trust**. NGOs depend on the trust of the public: the communities they serve, volunteers, and funders. If our AI-assisted decisions are found to be unfair or biased, public perception can quickly turn negative. A real-life example: an association in the US implemented a chatbot to guide people with eating disorders, but the chatbot ended up giving harmful advice due to flawed data, causing a scandal and forcing its removal. The lesson was clear: a *small ethical mistake* with AI can lead to a big reputational problem, eroding years of good image in the community.

There are also **legal risks**. If an NGO uses a third-party AI platform to, say, screen job applications or select beneficiaries, and that tool has discriminatory biases, we could face complaints or even lawsuits for discrimination. A lawyer specializing in the nonprofit sector warned that if there is bias in recruitment software and a person is not hired because of that bias, **the organization could face a legal case for discrimination.** In other words, the ethical and *legal* responsibility for automated decisions lies with us as users of the technology.

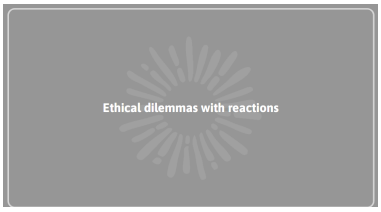
Finally, allowing bias affects our **effectiveness and impact**. Unfair decisions lead to the misallocation of scarce resources: we could focus on less needy populations while neglecting those we are trying to empower, perpetuating the inequality we wanted to combat. In short, uncorrected algorithmic bias is contrary to *the DNA* of NGOs (which seek

social justice) and can cause us to fail in our fundamental objectives.

Minute 10	Facilitator Action
	Invite the group to reflect.

Script/instructions

Do we know of any examples close to home (in our region or sector) where a digital tool has unintentionally discriminated? How would our organization react if something like this happened? Think about it for a couple of minutes!

Minute 11	Facilitator's action
	

Script/instructions

We sometimes believe that ethical dilemmas are abstract or technical, but in the work of NGOs, they take concrete form every day. Who prioritizes? How is it decided who receives help? Are we using tools that truly understand our context, or are we simply replicating inequalities disguised as efficiency?

With this activity, we seek to stimulate that reflection. There are no right or wrong answers, but there are decisions that we must make conscientiously, as a team, and with clear criteria.

So, as you respond with emojis, think about: Who could be affected by this decision? Whose voice is missing from this process? How could I make it fairer?

Let's review each case together and discuss it briefly. The goal is not to judge, but to learn how to improve our AI practices from an ethical, human, and critical perspective.

Minute 12

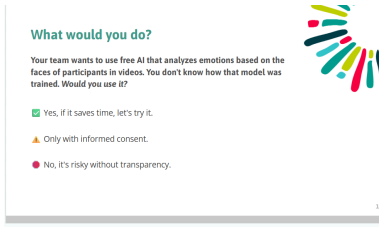
Facilitator's action

State the question and the situation and ask for reactions with emojis in the chat.

Script/instructions

Let's see your answers.

Minute 13



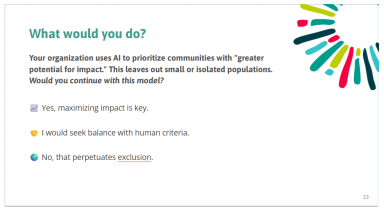
Facilitator's action

State the question and the situation and ask for reactions using emojis in the chat.

Script/instructions

Let's see your answers.

Minute 14



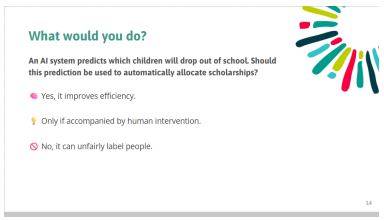
Facilitator Action

State the question and the situation and ask for reactions using emojis in the chat.

Script/instructions

Let's see your answers.

Minute 15

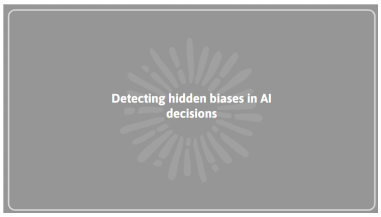


Facilitator Action

State the question and the situation and ask for reactions using emojis in the chat.

Script/instructions

Let's see your answers.

Minute 16	**Facilitator's Action**
	Read the title and continue.
Script/instructions Now, let's detect hidden biases in AI decisions	

Minute 17	**Facilitator Action**
	Introduce the activity, divide the groups randomly, and allow 10 minutes for discussion. Once everyone returns, debrief the activity.

Script/instructions

Let's do a little exercise to apply what we've just seen. Here is a scenario based on real events: *A nonprofit organization decides to streamline its hiring process by using an AI tool that ranks candidates based on their resumes. After a while, they notice a worrying pattern: most of the resumes recommended by the AI belong to men, and many qualified female candidates are left out of the shortlist.* This conflicts with the NGO's commitment to gender equality.

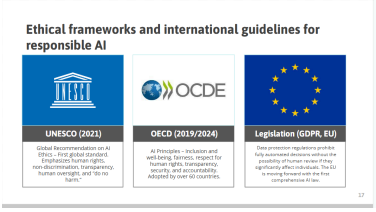
Let's divide the room into small groups. Each group will discuss the questions on the screen for about 5 minutes: **What kind of bias do you think is present in this system? How could you detect or confirm it? And most importantly: if you were the NGO team, what would you do to correct or mitigate this bias?**

Think about the possible causes: Could it be that the training data consisted mainly of resumes from men in previous positions? Perhaps the AI learned to value certain words or experiences that are more common in men due to cultural factors—just as happened in the Amazon case. Also, think about solutions: for example, reviewing the algorithm, incorporating diversity criteria into the selection process, or even removing certain variables (such as gender or gender proxies) to avoid this bias.

Debrief after the activity: This was a case of gender bias in AI. Possible solutions that may have emerged include: training the model with more balanced data, conducting blind tests where gender is hidden from the algorithm, or establishing human oversight to review AI recommendations before making decisions. The central idea is that we must audit and

adjust our AI tools to ensure that they do not betray our values. Thank you for your ideas; with this in mind, let's move on to how ethical frameworks help us prevent situations like this.

Minute 18



Facilitator Action

Script/instructions

Fortunately, we are not starting from scratch in responding to these challenges: in recent years, **international ethical frameworks** and **regulations** have emerged to guide us. We will highlight a few:

In 2021, **UNESCO** (the UN Educational, Scientific and Cultural Organization) succeeded in getting 193 member states—practically the entire world—to adopt a historic agreement: the *Recommendation on the Ethics of Artificial Intelligence*. It is the first global standard on the subject. What does this document say? Its *cornerstone* is **to protect human dignity and human rights** in all AI development. It sets out principles such as **transparency and fairness** in algorithms, the importance of **human oversight** (not delegating decisions entirely to machines), promoting diversity and inclusion, and ensuring that AI does not cause harm or deepen inequalities. Gabriela Ramos, Deputy Director-General of UNESCO, has clearly stated that governments and organizations must develop AI in a way that *does not reproduce the biases of the real world that we dislike,* thus building more **inclusive** societies **free of discrimination.** The UNESCO Recommendation is an **ethical beacon:** it is not a binding law, but it guides countries (and therefore the institutions within them) on how to establish good practices and policies in AI, from data management to education and research in this field.

On the other hand, the **OECD (Organization for Economic Cooperation and Development)** published *Principles for AI* in 2019, which quickly gained international support (G20, EU, US, etc.). These principles focus on shared values: promoting **inclusive growth** and general well-being through AI, **respecting human rights, democracy, and equity,** ensuring the **transparency** and **explainability** of systems, guaranteeing the **robustness, security, and absence of bias** in algorithms, and ensuring the accountability of those who develop and use AI. In 2024, the OECD updated these principles to reinforce issues such as **information integrity** (avoiding AI-generated misinformation) and **security** with control mechanisms over potentially harmful systems. For NGOs, these OECD values provide a common language for our policies: for example, we align ourselves with the principle of transparency if we explain to our beneficiaries how an AI made a certain

decision, or with the principle of inclusion if we ensure that the technology benefits all groups and not just a few.

In addition to ethical frameworks, there are already **laws and regulations** in place. The best known is the **GDPR** (General Data Protection Regulation of the European Union), in force since 2018, which, although focused on privacy, has an article (22) that is very relevant to AI: **it gives people the right not to be subject to purely automated decisions that have significant effects on them, without human intervention**. In other words, in Europe (and de facto for any entity that processes European data, including international NGOs), an algorithm could not be used to, say, accept or reject someone in a program **without** a human review mechanism. This legal protection seeks to curb "power imbalances" between individuals and opaque algorithms, and essentially forces the introduction **of human oversight and transparency,** which is key to mitigating bias.

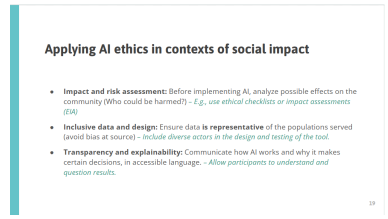
The European Union is also discussing the **European AI Act**, which will be the first comprehensive law specifically on artificial intelligence worldwide. It is not yet in force (final approval is expected soon), but it is already known that it will classify AI systems by risk level (high, medium, low) and impose strict requirements—such as compliance assessments, documentation, and bias elimination—on those uses considered *high risk* (e.g., in health, employment, public safety). This means that AI providers and user organizations will have obligations to ensure the **ethics and safety** of their systems. To illustrate its relevance: the European Parliament has already put forward a draft, and it is known as the first comprehensive regulation on the subject.

Outside Europe, other countries are following suit: for example, the United States published a **Blueprint for an AI Bill of Rights** with principles for developers (such as preventing algorithmic discrimination and ensuring explainability), and several Latin American countries are adopting or discussing national AI strategies with ethical components. Even at the United Nations level, there is talk of a possible global instrument in the future.

Minute 19	Facilitator Action
	

Script/instructions

In summary, NGOs are not alone in this ethical conversation: we can base our internal policies on solid references such as UNESCO and the OECD, and we must be aware of the laws in force (e.g., data protection) that apply to us locally. These guidelines help us translate values into concrete actions when using AI.

Minute 20	Facilitator Action
	

Script/instructions

Let's move from theory to practice: How do we apply these ethical principles in our daily operations and AI projects? Here are some **concrete actions** that an NGO can take:

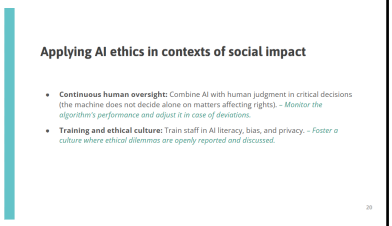
- **Impact and risk assessment:** Before launching an AI tool, let's pause and assess*: What could go wrong?* For example, if we are going to use an algorithm to decide on the allocation of microcredits in a community, we must analyze whether it could, by its design, exclude any group (women, ethnic minorities, people over a certain age, etc.). A good habit is to conduct an **Ethical Impact Assessment** (inspired by UNESCO's proposal) where we identify risks of bias, invasion of privacy, or other harms, and plan how to mitigate them. There are checklists or even formalized methodologies for this—the idea is **not to rush into using AI blindly,** but to anticipate the social and ethical consequences.

- **Inclusive data use and design:** Much of the bias comes from the data, so that is where we must act first. Let's make sure that the data we feed into AI **represents the diversity** of the people we serve. If we are going to create a model for a regional initiative, for example, let's incorporate data from all areas, not just the capital. And if we notice that we are missing data from a certain group (let's say we have fewer cases of rural women in the database), let's be aware of that gap. Sometimes the solution is to supplement with external data or even **manually adjust** certain parameters to correct the bias. In addition, let's involve stakeholders in the design: why not invite community members or human rights experts to test the pilot system and provide feedback? That diverse feedback can reveal biases that we technicians overlook.

- **Transparency and explainability:** To gain the trust of our beneficiaries and staff, we must be transparent about the use of AI. This involves two things: first, **disclosing** that AI was used in the decision (for example, "the scores were generated with a predictive model based on X criteria"), and second, **explaining in simple terms** the reasons for the decision. No matter how complex the algorithm is, the person affected deserves to know the basic reasons: "We are offering you this assistance because the system identified that you meet certain priority criteria (A, B, C)." Explainability is also internal: the technical team must be able to interpret what factors the AI is using to make decisions. If it is an incomprehensible

"black box," it is difficult to trust it or detect biases. Transparency creates an environment of accountability and allows anyone to question and appeal decisions if something seems wrong.

Minute 21



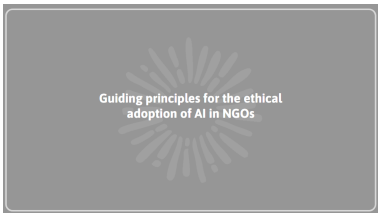
Facilitator Action

Script/instructions

- **Continuous human oversight:** A fundamental principle is **not to relinquish absolute control**. No high-impact decision in our NGO should be made 100% by a machine without someone validating it. For example, if an algorithm suggests excluding a family from a program because it considers them "unqualified," a human worker must review that case before the final decision, and can reverse it if they see something unfair. AI should be *like an assistant,* not the director. In addition, its performance must **be monitored** over time: indicators of possible biases must be established (are we reaching different demographic groups in a balanced way? Have there been complaints of unfair decisions?) and the results must be audited periodically. If we detect deviations, we should adjust or retrain the system. AI ethics is not "set and forget"; it requires constant monitoring, just like any NGO program.

- **Training and ethical culture:** Ultimately, none of this works if our team is not aware of the issues. It is essential to train our staff—both technicians and those who implement the programs—in basic AI concepts, in understanding what algorithmic bias is, what the legislation says (for example, GDPR on personal data protection), how to handle data responsibly, etc. But beyond knowledge, we must cultivate an **organizational culture** where ethical dilemmas are taken seriously. Encourage anyone on the team who notices a possible bias or inappropriate use of data to speak up without fear. Perhaps that means creating a small digital ethics committee or including this topic in project meetings. The important thing is that AI ethics becomes an ongoing conversation within the NGO, not a one-time check.

By applying these practices, NGOs can **leverage AI while minimizing its risks**. As a report by Nonprofit Quarterly mentions, having clear policies and guardrails not only prevents harm, but also allows us to use AI **ethically and consciously to empower our social mission** rather than undermine it.

Minute 22	Facilitator Action
	Read the title and continue

Script/instructions

Guiding principles for the ethical adoption of AI in NGOs

Minute 23	Facilitator's action
	Explain each of the proposed principles one by one.

Script/instructions

We now propose a list of **Guiding Principles** that could form the "code of ethics" for the use of AI in an NGO. These principles act as the backbone for any technological policy or decision we make:

1. **Equity and Inclusion:** Any AI solution we adopt must be examined under the lens of equity. Is it accessible to different groups? Are its results fair to women, men, ethnic minorities, people with disabilities, different ages, and geographic areas? We are committed to identifying and eliminating any bias that leads to unequal treatment. UNESCO highlights this point in its recommendation: AI must promote social justice and **non-discrimination,** ensuring that its benefits reach everyone. For example, if we develop a customer service bot, we will verify that it understands and includes the cultural and linguistic diversity of our users, avoiding stereotypical responses.

2. **Transparency and Explainability:** As a matter of principle, we will not use "black boxes" irresponsibly. We will be transparent about when and how we use AI. If a donor receives an AI-generated email or a beneficiary is evaluated by an algorithm, we will communicate this openly. And if someone asks, "Why did this algorithm decide this way?", we will be able to explain it in simple language. Transparency builds trust and allows third parties to scrutinize and point out potential flaws. Imagine that a beneficiary questions why they were not selected for a program; under this principle, we must be able to review the decision, explain it ("we

prioritize cases with such characteristics because...") and correct it if we find an error. No "the computer said so and it is infallible."

3. **Responsibility and Accountability:** Even if we use third-party AI or automation, *the ultimate responsibility is ours*. This principle establishes that there will be **internal accountability** for the ethical functioning of AI. How does this look? On the one hand, by clearly assigning roles: who supervises the algorithm, who to report incidents to, who decides on adjustments. On the other hand, by ensuring *auditability:* recording how automated decisions are made so that we can reconstruct and understand a failure. If, for example, a family was mistakenly excluded from receiving aid, we must investigate what happened in the algorithmic process and correct it. It also involves creating channels for users or employees to report biases or problematic results without bureaucracy. In line with international frameworks, we will promote **algorithms that are auditable, traceable, and subject to proper control.** The NGO will be accountable for its tools just as it is accountable for the actions of its staff.
4. **Privacy and Data Security:** This principle reinforces our duty to protect personal data to the highest degree. We will only use people's data if we have clear authorization and a legitimate purpose. If we use AI that requires sensitive data (health, socioeconomic status), we will ensure compliance with regulations such as the GDPR and local regulations, minimizing collection to the minimum necessary. In addition, we will implement **cybersecurity** measures to prevent breaches or unauthorized access to data or models. Ethics also means not exposing those who have entrusted their data to us. For example, if we feed an algorithm with information about our beneficiaries, we will never do so on public platforms or without reviewing the terms and conditions (preventing that data from being misused by third parties). The confidentiality and integrity of information are sacred.
5. **Human Oversight:** This is key: **no AI will operate on absolute autopilot**, especially in decisions that affect people. There will always be a human being in the decision-making loop to validate or have the final say. This principle avoids many risks. Suppose a system recommends rejecting someone from a program; with human oversight, that person implicitly has the right to a second human review or appeal. In addition, we maintain control: AI assists us but *does not replace our human responsibility.* As specific policies, we could say: "if an algorithm flags a case as fraudulent, a human auditor will review it before taking action" or "the NGO's chatbot responses will be regularly monitored by a human supervisor to correct any inappropriate responses." This also aligns with legislation such as Article 22 of the GDPR that we mentioned: in our NGOs, we will always comply by providing the option of human intervention.

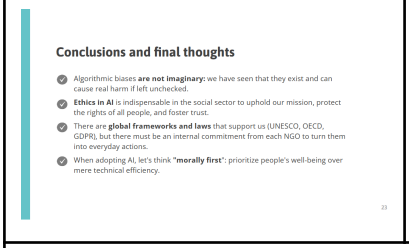
6. **Proportionality and Do No Harm:** Inspired by bioethics and adopted by UNESCO, this principle invites us to ask ourselves: *Do we really need to use AI for this?* And if we do use it, have we calibrated its scope well so as not to cause accidental harm? **Proportionality** means that the complexity or intrusiveness of AI must be commensurate with the problem it solves. For example, we would not

justify implementing facial recognition (with all its privacy and racial bias risks) just to take attendance at a training session, right? But perhaps we would for a critical family reunification project where there is no other way to identify people. "Do no harm" is a reminder of our humanitarian mandate: if AI presents a high possibility of harming our beneficiaries (e.g., labeling them with scores that could stigmatize them), we will choose not to use it. We will always weigh the benefits against the potential harms before adopting the technology.

7. **Training and Continuous Improvement:** Finally, an internal principle: we are committed to **continuous learning and improvement** in this area. AI ethics are evolving, new challenges are emerging (think generative AI, deepfakes, etc.), and regulations are also being updated. Our guidelines cannot be static. So, we will include regular training for staff on data/AI, update our technology policies regularly, and share lessons learned with the community. It is also valid to join "AI for Good" networks or initiatives where NGOs share ethical practices. The humility to adjust course is key: we may implement a system with the best of intentions and then discover a bias; we must be ready to recognize it, report it to stakeholders, and correct it.

These guiding principles serve as a compass. Whenever we consider a new AI project or evaluate one that is already underway, we can check it against this list: Are we being fair? Transparent? Who is responsible? Etc. If any of the answers make us uncomfortable, it is a sign that we need to rethink the tool or the conditions of its use. The ethical adoption of AI is not a final state, but **a continuous process of conscious decision-making** aligned with our values as a social organization.

Minute 24



Facilitator Action

Script/instructions

To close this module, let's recap the key points that I hope you will take away with you:

- **Recognize the problem:** Algorithmic bias is real and documented. It is not a theoretical fear or something that "only happens to others." As we have seen, from health to public safety to human resources, algorithms can exacerbate pre-existing inequalities if we are not careful. As NGOs, we must start from the premise that *yes, algorithms can be wrong and even unfair,* and therefore require our constant

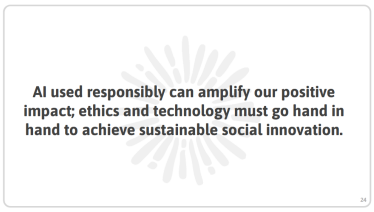
ethical vigilance. We cannot blindly delegate decisions to AI.

- **Ethics = Essential in our organizations:** We have discussed how our core values (equality, justice, inclusion, non-discrimination) must be reflected in any use of advanced technology. AI ethics is not a "gimmick" or a compliment; it is as important as our codes of conduct or our beneficiary protection safeguards. If we neglect ethics, we risk causing harm to those we want to help and losing the trust we have worked so hard to earn in communities and among allies. Conversely, if we incorporate ethics from the design stage, AI can be a powerful ally for social good.

- **Rely on existing guidelines and standards:** We are not alone on this journey. We have valuable international guidance—for example, we can align our policies with the **OECD Principles,** which address transparency, fairness, accountability, etc., or with the **UNESCO Recommendation,** which, let us remember, emphasizes the primacy of human rights and the need to avoid repeating the vices of our society with AI. Likewise, current legislation such as the GDPR provides us with a legal framework to follow, especially in terms of data protection and automated decision-making. Keeping up to date with these references gives us legitimacy and solidity in our practices.

- **Internal and daily commitment:** Nothing changes if these issues remain only on paper. The ethical adoption of AI requires a commitment from *the entire organization*, from senior management—which must promote and provide resources to implement these policies—to every technician or user who interacts with an intelligent system. It is a cross-cutting effort: for example, the Human Resources department evaluating the fairness of a selection tool, the programs department verifying that a targeting system is not biased, the communications department being transparent about how we use chatbots, etc. In other words, there must be a culture of "**thinking first about the human and ethical, then about the** digital" in every decision we make with AI.

- **The positive side:** I want to make it very clear that this module is not intended to scare you into giving up on AI. On the contrary, when used well, AI can **greatly enhance social impact**. It can help us analyze needs data faster, personalize interventions, free up administrative time to focus on community work, and even innovate solutions to complex problems (we saw examples of NGOs using AI to monitor climate change, improve agriculture, etc.). The key is to "take ethical precautions." As a recent article pointed out, if we first address the potential ethical and legal risks, we can use AI to *maximize social benefits,* providing a positive counterbalance to the use of AI in other sectors focused solely on profit.

Minute 25	Facilitator Action
	

Script/instructions

In short, **ethics and bias in AI** are not a passing issue; they are here to stay. But with knowledge, clear principles, and proactive action, our NGOs can lead by example in the responsible adoption of AI. In this way, we will demonstrate that the most advanced technology **can indeed** serve humanity and justice, if implemented conscientiously. I invite you to take these conversations back to your teams, review your projects in light of what you have learned, and, above all, remain curious and critical about AI. Only then can we navigate this new terrain together, taking advantage of its benefits without losing sight of our ethical compass.

Thank you all for your active participation! I look forward to any final questions or comments you may wish to share.